

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:50:35 ; Search time 48.0508 Seconds
(without alignments)
124.790 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 229

Sequence: 1 QPLRRQVVVVPLSLVLL.....RLPLKRMGGSYRCVNASTAN 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.5	26.0	43	AA005294	Human polypeptide
2	59.5	26.0	1148	AAU36302	Pseudomonas aerugi
3	59	25.8	109	AAU96656	Human reproductive
4	59	25.8	116	AAU93273	Human protein sequ
5	57.5	25.1	732	AAU42697	Mouse serine-threo
6	57.5	25.1	733	AAU42696	Rat serine-threoni
7	57.5	25.1	806	AAU42692	Novel protein kina
8	56.5	24.7	415	AAU56275	Arabidopsis thalia
9	56	24.5	89	ABG04156	Novel human diagn
10	55	24.0	70	AA012427	Human polypeptide

11	55	24.0	497	22	ABG30010	Novel human diagn
12	55	24.0	642	20	AAU01541	Alpha(2-3) trans-s
13	55	24.0	669	23	ABU08420	Catalytic trans-si
14	55	24.0	1060	20	AAU01540	Trypanosoma cruzi
15	53	23.1	64	22	AAU05255	Mouse Nope (neighb
16	53	23.1	68	22	AAU02747	Human polypeptide
17	53	23.1	932	22	AAU05252	Mouse Nope (neighb
18	53	23.1	1252	22	AAU05251	Propionibacterium
19	52.5	22.9	87	22	AAU1536	Mouse Nope (neighb
20	52.5	22.9	158	18	AAU11790	Early onset Alzhei
21	52.5	22.9	318	22	AAU09056	C glutamicum prote
22	52	22.7	141	21	AAU58870	Arabidopsis thalia
23	52	22.7	146	21	AAU58869	Arabidopsis thalia
24	52	22.7	1023	22	ABU70491	Drosophila melanog
25	51	22.3	112	22	AAU52745	Propionibacterium
26	51	22.3	173	22	ABU15535	Novel human diagn
27	51	22.3	236	22	ABU14569	Novel human diagn
28	50.5	22.1	177	20	AAU43529	Rabbit interleukin
29	50.5	22.1	177	22	AAU66667	Rabbit interleukin
30	50	21.8	44	20	AAU35913	Human prostate tum
31	50	21.8	149	20	AAU74068	Lolium perenne LpP
32	50	21.8	246	23	ABU78981	Human prostate tum
33	50	21.8	264	22	AAU40175	Human prostate tum
34	50	21.8	411	22	ABU08428	Propionibacterium
35	50	21.8	3138	22	ABU61958	Novel human diagn
36	49.5	21.6	165	22	AAU75631	Drosophila melanog
37	49.5	21.6	288	21	AAU24977	Human colon cancer
38	49.5	21.6	786	22	ABU29559	Arabidopsis thalia
39	49.5	21.6	3854	22	ABU67104	Novel human diagn
40	49.5	21.6	5385	22	ABU66487	Drosophila melanog
41	49.5	21.6	5496	22	ABU67161	Drosophila melanog
42	49.5	21.6	8805	22	AAU67112	Drosophila melanog
43	49	21.4	95	22	AAU86372	Human immune/haema
44	49	21.4	182	22	AAU14311	Human novel protei
45	49	21.4	336	22	AAU52192	Propionibacterium

ALIGNMENTS

RESULT 1

AA005294
ID AA005294 standard; Protein; 43 AA.

XX AA005294;

XX AC (first entry)

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 19186.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AA185225.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -

PS Claim 20; SEQ ID NO 19186; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 43 AA;

Query Match 26.0%; Score 59.5; DB 22; Length 43;
Best Local Similarity 48.6%; Pred. No. 0.13;
Matches 17; Conservative 2; Mismatches 7; Indels 9; Gaps 2;

QY 13 LSPRL-----VLLAFCRORLPKMGSGSYRCVNAS 42

Db 4 LSPRLWCSGTILAHCSURLP-----GSSASCVSAS 34

RESULT 2

AAU36302
ID AAU36302 standard; Protein; 1148 AA.

XX AC AAU36302;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #292.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253623P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS54161.

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11895; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1148 AA;

Query Match 26.0%; Score 59.5; DB 22; Length 1148;
Best Local Similarity 33.9%; Pred. No. 5.5;
Matches 19; Conservative 7; Mismatches 13; Indels 17; Gaps 2;

QY 2 PLRRQRVVVVPLSPRLVLLAFCR-----QRLPLKRM-----GGSYRCVN 40

Db 99 PQKHGVLVVPISTALHRLAPTRFLIGSSILVDVGQKLDVERMLRLEGAGYRCVD 154

RESULT 3

AAU96656

ID AAU96656 standard; Protein; 109 AA.

XX AC AAU96656;

XX 21-NOV-2001 (first entry)

XX Human reproductive system related antigen SEQ ID NO: 5314.

XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy.

XX Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 28-JUN-2000; 2000US-0209467.

XX 30-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216847.

XX 11-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0217496.

XX 26-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

AAB93273
 ID AAB93273 standard; Protein; 116 AA.
 AC AAB93273;
 XX
 XX
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:12312.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 OS EP1074617-A2.
 PN 07-FEB-2001.
 PD
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 FI
 XX WPI; 2001-318749/34.
 DR
 XX
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 12312; 2537pp + CD ROM; English.
 PS
 XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 116 AA;
 Query Match 25.8%; Score 59; DB 22; Length 116;
 Best Local Similarity 35.6%; Pred. No. 0.49;
 Matches 16; Conservative 2; Mismatches 15; Indels 12; Gaps 2;
 QY 13 LSPRL-----VLLAFRCQRLP-----LKMGGSYRCVNSTAN 45
 DB 5 LLPRLECSGTLVHCNLRPLGSSDLSASASRVAGDYRCATPCLAN 49

RESULT 5
 AAY42697
 ID AAY42697 standard; Protein; 732 AA.
 XX
 AC AAY42697;
 XX
 XX
 DT 17-JAN-2000 (first entry)
 DE Mouse serine-threonine protein kinase PK80 sequence.
 XX Serine-threonine protein kinase; PK80; kidney; drug; tumor; mouse;
 KW kidney disease.
 XX
 XX Mus musculus.
 OS WO9950395-A1.
 PN 07-OCT-1999.
 PD
 XX 26-MAR-1999; 99WO-JP01575.
 PR 27-MAR-1998; 98JP-0100261.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Muramatsu M, Tokumitsu H, Noguchi T;
 PI
 XX WPI; 1999-591314/50.
 DR N-PSDB; AAZ23071.
 XX
 XX Serine-threonine protein kinase PK80, useful for screening potential
 PT PK80 regulators, useful for treating kidney disease -
 PT
 XX Claim 3; Page 41-51; 64pp; Japanese.
 PS
 XX The invention provides serine-threonine protein kinases (designated PK80)
 CC expressed in the kidneys of rats and mice. The rat or mouse PK80 protein
 CC can be recombinantly expressed by standard recombinant methodology. It is
 CC used for screening of potential PK80 regulatory compounds to obtain drugs
 CC for the prevention and treatment of diseases with which PK80 is
 CC associated, e.g. tumors and other diseases of the kidney. The present
 CC sequence represents the amino acid sequence of the mouse PK80 protein.
 XX
 SQ Sequence 732 AA;
 Query Match 25.1%; Score 57.5; DB 20; Length 732;
 Best Local Similarity 48.5%; Pred. No. 6.7;
 Matches 16; Conservative 3; Mismatches 7; Indels 7; Gaps 2;
 QY 11 VPLSPRLVLLAFRCQRLPDKRMGGSYRCVNSTAN 43
 DB 130 VP-SPRL-----RSRLPSKLLKSGSHRCGEAGS 155
 RESULT 6
 AAY42696
 ID AAY42696 standard; Protein; 733 AA.
 XX
 AC AAY42696;
 XX
 DT 17-JAN-2000 (first entry)
 DE Rat serine-threonine protein kinase PK80 sequence.
 XX Serine-threonine protein kinase; PK80; kidney; drug; tumor; rat;
 KW kidney disease.
 XX
 XX Rattus norvegicus.
 OS WO9950395-A1.
 PN 07-OCT-1999.
 PD
 XX

PF 26-MAR-1999; 99WO-JP01575.
 PR 27-MAR-1998; 98JP-0100261.
 XX (HELI-) HELIX RES INST.
 XX Muramatsu M, Tokumitsu H, Noguchi T;
 XX WPI; 1999-591314/50.
 DR N-PSDB; AA223070.
 XX Serine-threonine protein kinase PK80, useful for screening potential
 PT PK80 regulators, useful for treating kidney disease -
 XX Claim 1; Page 24-34; 64pp; Japanese.
 XX The invention provides serine-threonine protein kinases (designated PK80)
 CC expressed in the kidneys of rats and mice. The rat or mouse PK80 protein
 CC can be recombinantly expressed by standard recombinant methodology. It is
 CC used for screening of potential PK80 regulatory compounds to obtain drugs
 CC for the prevention and treatment of diseases with which PK80 is
 CC associated, e.g. tumors and other diseases of the kidney. The present
 CC sequence represents the amino acid sequence of the rat PK80 protein.
 XX Sequence 733 AA;
 SQ Query Match 25.1%; Score 57.5; DB 20; Length 733;
 Best Local Similarity 48.5%; Pred. No. 6.7;
 Matches 16; Conservative 3; Mismatches 7; Indels 7; Gaps 2;
 QY 11 VPLSPRLVLLAFRCORLPKRMGGSYRCVNST 43
 DB 130 VP-SPRL-----RSRLPSKLLKGSHRCGEAGS 155
 RESULT 7
 AAB65622
 ID AAB65622 standard; Protein; 806 AA.
 XX AAB65622;
 XX 27-MAR-2001 (first entry)
 DT Novel protein kinase, SEQ ID NO: 148.
 DE Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 XX immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
 XX Mus musculus.
 OS WO2000073469-A2.
 PN 07-DEC-2000.
 PD 26-MAY-2000; 2000WO-US14842.
 XX 28-MAY-1999; 99US-0136503.
 XX (SUGE-) SUGEN INC.
 XX Plowman GD, Martinez R, Whyte D, Sudersanam S;
 PI WPI; 2001-032161/04.
 DR N-PSDB; AAF44648.
 XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 XX

PS The present sequence is a novel protein kinase. The novel protein kinases
 XX and the nucleic acids that encode them may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 XX reproductive disorders.
 SQ Sequence 806 AA;
 Query Match 25.1%; Score 57.5; DB 22; Length 806;
 Best Local Similarity 48.5%; Pred. No. 7.5;
 Matches 16; Conservative 3; Mismatches 7; Indels 7; Gaps 2;
 QY 11 VPLSPRLVLLAFRCORLPKRMGGSYRCVNST 43
 DB 204 VP-SPRL-----RSRLPSKLLKGSHRCGEAGS 229
 RESULT 8
 AAG56275
 ID AAG56275 standard; Protein; 415 AA.
 XX AAG56275;
 XX 18-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 72310.
 DE Arabidopsis thaliana.
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS EP1033405-A2.
 XX 06-SEP-2000.
 PD 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.

PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	27-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139750.	PR	30-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139750.	PR	30-AUG-1999;	99US-0151303.
PR	21-JUN-1999;	99US-0139817.	PR	31-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0139899.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140033.	PR	07-SEP-1999;	99US-0152363.
PR	23-JUN-1999;	99US-0140354.	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	99US-0140695.	PR	13-SEP-1999;	99US-0153758.
PR	24-JUN-1999;	99US-0140823.	PR	15-SEP-1999;	99US-0154018.
PR	28-JUN-1999;	99US-0140981.	PR	16-SEP-1999;	99US-0154039.
PR	30-JUN-1999;	99US-0141287.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0141842.	PR	22-SEP-1999;	99US-0155139.
PR	01-JUL-1999;	99US-0142154.	PR	23-SEP-1999;	99US-0155486.
PR	02-JUL-1999;	99US-0142055.	PR	24-SEP-1999;	99US-0155659.
PR	06-JUL-1999;	99US-0142330.	PR	28-SEP-1999;	99US-0156458.
PR	08-JUL-1999;	99US-0142803.	PR	29-SEP-1999;	99US-0156596.
PR	09-JUL-1999;	99US-0142920.	PR	04-OCT-1999;	99US-0157117.
PR	12-JUL-1999;	99US-0142977.	PR	05-OCT-1999;	99US-0157753.
PR	13-JUL-1999;	99US-0143354.	PR	06-OCT-1999;	99US-0157865.
PR	14-JUL-1999;	99US-0143624.	PR	07-OCT-1999;	99US-0158029.
PR	15-JUL-1999;	99US-0144005.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144085.	PR	12-OCT-1999;	99US-0158369.
PR	16-JUL-1999;	99US-0144086.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159331.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159337.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144632.	PR	18-OCT-1999;	99US-0159584.
PR	20-JUL-1999;	99US-0144884.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160768.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR		

PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 24.7%; Score 56.5; DB 21; Length 415;
 Best Local Similarity 42.9%; Pred. No. 5;
 Matches 18; Conservative 5; Mismatches 14; Indels 5; Gaps 3;
 QY 7 RVVVVPLSRLLVLLAFRCR-QLRPLKR--MGGSYRC--VNAST 43
 ||: ||| : ||||| : ||| : |||
 Db 14 RVLIFPCSPRSSFVFTSRLLSLPLKRASIGCAVSCGVNGLT 55
 ||: ||| : ||||| : ||| : |||
 RESULT 9
 ABG04156
 ID ABG04156 standard; Protein; 89 AA.
 XX
 AC ABG04156;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #4147.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS68343.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 34515; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. ABG00010-ABG30377 represent novel human
 diagnostic amino acid sequences of the invention.
 Note: The sequence data for this patent did not appear in the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 89 AA;
 Query Match 24.5%; Score 56; DB 22; Length 89;
 Best Local Similarity 36.8%; Pred. No. 1;
 Matches 14; Conservative 6; Mismatches 14; Indels 4; Gaps 1;
 QY 4 RRRVVVPLSRLLVLLAFRCRQLRPLKRMGGSYRCVNA 41
 ||: ||| : ||| : ||| : ||| : |||
 Db 42 RRRDPPRAPAPR----PWCQRAALSSLGSHLCDDA 75
 ||: ||| : ||| : ||| : ||| : |||
 RESULT 10
 AAO12427
 ID AAO12427 standard; Protein; 70 AA.
 XX
 AC AAO12427;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 26319.
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI92358.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 26319; 1399pp + Sequence Listing; English.
 CC
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed


```

Db      436 AYRCVNASTAN 446
      :|||||
RESULT 13
ID      ABB08420 standard; Protein; 669 AA.
XX
AC      ABB08420;
XX
DT      01-JUL-2002 (first entry)
XX
DE      Catalytic trans-sialidase unit of T. cruzi amino acid sequence.
XX
KW      Mycoplasma associated disease; cell proliferation; trans-sialidase;
KW      enzyme; atherosclerotic vascular disease; malignancy;
KW      sialic acid; antiatherosclerotic; antibacterial; antiviral; anti-HIV;
KW      cytostatic; vasotropic; ovarian carcinoma; breast cancer;
KW      prostate cancer; colon cancer; lung cancer; leukaemia; HIV;
KW      human immunodeficiency virus; chlamydia; PCR primer.
XX
OS      Trypanosoma cruzi.
OS      Synthetic.
XX
PN      WO200202050-A2.
XX
PD      10-JAN-2002.
XX
PF      03-JUL-2001; 2001WO-BR00083.
XX
PR      03-JUL-2000; 2000BR-0002989.
XX
PA      (HIGUCHI) HIGUCHI M D L.
PA      (SCHE/) SCHENKMAN S.
XX
PI      Higuchi MDL, Schenkman S;
XX
WPI; 2002-154675/20.
DR      N-PSDB; ABA98876.
XX
Composition useful for treatment of mycoplasma infection and diseases
PT      associated with cell proliferation e.g. malignancy or with co-infection
PT      with another microbe, comprises agent inhibiting sialic acid-mediated
PT      attachment of mycoplasma -
XX
Claim 6; Fig 26; 63pp; English.
XX
The invention relates to a composition useful for treating or preventing
CC      mycoplasma infection in a subject suffering from a disorder characterised
CC      by increased cell proliferation or by co-infection with a second microbe,
CC      comprising an agent that prevents or inhibits sialic acid-mediated
CC      attachment of mycoplasma to the subject's cells. The activity of
CC      compositions of the invention may be described as; antiatherosclerotic,
CC      antibacterial, antiviral, anti-HIV, cytostatic and vasotropic. The
CC      compositions are useful to treat diseases associated with undesirable
CC      cell proliferation, such as atherosclerotic vascular disease and
CC      malignancy, by reducing or preventing mycoplasma infection. Examples of
CC      malignancies include; ovarian carcinoma, breast cancer, prostate cancer,
CC      colon cancer, lung cancer and leukaemia. They are also useful to treat
CC      diseases associated with infection with other infectious organisms
CC      co-occurring with mycoplasma (and typically increasing the virulence of
CC      both pathogens), especially human immunodeficiency virus or chlamydia
CC      species. They can be used to treat such diseases in humans or other
CC      animals, and can be administered in conjunction with conventional agents
CC      e.g. anti-platelet or chemotherapeutic agents. The current sequence
CC      represents the catalytic trans-sialidase unit of T. cruzi amino acid
CC      sequence.
XX
Sequence 669 AA;
Query Match 24.0%; Score 55; DB 23; Length 669;
Best Local Similarity 90.9%; Pred. No. 15;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      35 SYRCVNASTAN 45
      :|||||
Db      455 AYRCVNASTAN 465
      :|||||
RESULT 14
ID      AAY01540 standard; Protein; 1060 AA.
XX
AC      AAY01540;
XX
DT      15-JUN-1999 (first entry)
XX
DE      Trypanosoma cruzi alpha (2-3) trans-sialidase amino acid sequence.
XX
KW      Alpha (2-3) trans-sialidase; sialyl-oligosaccharide; sialyllactose;
KW      cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW      duodenal ulcer; arthritis; enterotoxin.
XX
OS      Trypanosoma cruzi.
XX
PN      WO9908511-A1.
XX
PD      25-FEB-1999.
XX
PF      13-AUG-1998; 98WO-US16756.
XX
PR      14-AUG-1997; 97US-0911393.
XX
PA      (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI      Barker WA, Hakes DJ, Pelletier M, Zopf DA;
XX
WPI; 1999-190079/16.
DR      N-PSDB; AAX26611.
XX
Production of sialyl-oligosaccharides, particularly sialyl-lactose
PT      - by treating a dairy source such as a cheese processing waste
PT      stream with an alpha (2-3) trans-sialidase
XX
Disclosure; Fig 2; 84pp; English.
XX
The present sequence represents Trypanosoma cruzi alpha (2-3) trans-
CC      sialidase. The protein is used in the method of the invention to
CC      produce sialyl-oligosaccharides, particularly sialyllactose, which
CC      are produced by treating a dairy source such as a cheese processing
CC      waste strain with an alpha (2-3) trans-sialidase. The method can be
CC      used for producing sialyl-oligosaccharides, such as (2-3)sialyllactose
CC      for pharmaceutical use. (2-3)sialyllactose has been shown to neutralise
CC      enterotoxins of various pathogenic microbes including E. coli, Vibrio
CC      cholerae and Salmonella. It has also been shown that
CC      alpha (2-3) (2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc)
CC      interferes with colonisation of Helicobacter pylori and thereby prevents
CC      or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has
CC      additionally been proposed to inhibit immune complex formation by
CC      disrupting occupancy of the Fc carbohydrate binding site on IgG and to
CC      be useful in treating arthritis.
XX
Sequence 1060 AA;
Query Match 24.0%; Score 55; DB 20; Length 1060;
Best Local Similarity 90.9%; Pred. No. 25;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      35 SYRCVNASTAN 45
      :|||||
Db      467 AYRCVNASTAN 477
      :|||||
RESULT 15
ID      AAE05255 standard; peptide; 64 AA.

```

XX AAE05255;
AC
XX
DT 12-SEP-2001 (first entry)
DE
XX Mouse Nope (neighbour of punc ell) immunoglobulin domain 2 (Ig2).
XX
KW Mouse; Nope; neighbour of punc ell; cytostatic; neuroprotective; vaccine;
KW gene therapy; cerebroprotective; colonic cancer; mental retardation;
KW tumour suppressor; chromosome 9; transgenic animal; genetic disorder;
KW obesity; Bardet-Biedl syndrome; autosomal recessive disorder; retinitis;
KW polydactyly; pigmentosa; hypogonadism; hypertension; diabetes mellitus;
KW renal anomaly; cardiovascular anomaly; immunoglobulin domain 2; Ig2.
XX
OS Mus musculus.
XX
XX WO200149714-A2.
PN
XX
PD 12-JUL-2001.
XX
XX 26-OCT-2000; 2000WO-US29698.
PF
XX
XX 04-JAN-2000; 2000US-0174496.
PR
XX 19-MAY-2000; 2000US-0205789.
PR
XX (NEUR-) NEUROSCIENCES RES FOUND INC.
PA
XX Salbaum JM;
PI
XX WPI; 2001-441846/47.
XX
DR N-PSDB; AAD10025.
DR
XX Murine Nope polypeptides and nucleic acids useful for preventing,
PT diagnosing and treating colonic cancer and Bardet-Biedl syndrome -
PT
XX Claim 3; Page 87; 99pp; English.
PS
XX The present invention relates to Nope (neighbour of punc ell) which is
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate Nope expression such as cancers especially colonic
CC cancer and genetic disorders, as Nope is thought to be a tumour
CC suppressor. Nope gene is located on chromosome 9 and is used in gene
CC therapy. Nope is used as vaccine. Nope gene may be administered to treat
CC diseases by rectifying mutations or deletions in a patient's genome that
CC affect the activity of Nope by expressing inactive proteins or to
CC supplement the patients own production of Nope polypeptides. Nope gene
CC is used to study the expression and function of Nope polypeptides and
CC their role in metabolism through the creation of transgenic animal
CC models. The anti-Nope antibodies and Nope antagonists may also be used
CC to down regulate Nope expression and activity for the treatment of
CC Bardet-Biedl syndrome which is an autosomal recessive disorder
CC characterised by mental retardation, obesity, polydactyly, retinitis
CC pigmentosa and hypogonadism. Patients with Bardet-Biedl syndrome have a
CC high incidence of hypertension, diabetes mellitus and renal and
CC cardiovascular anomalies. The present sequence is mouse Nope (neighbour
CC of punc ell) immunoglobulin domain 2 (Ig2).
XX
SQ Sequence 64 AA;
Query Match 23.1%; Score 53; DB 22; Length 64;
Best Local Similarity 42.1%; Pred. No. 2;
Matches 16; Conservative 6; Mismatches 14; Indels 2; Gaps 2;
Qy 9 VVPLSPRLVLL-AFCRORLPLKRM-GGSYRCVNSTA 44
Db 26 VTPPEPRLLTLPKWLQILDVQSDAGSYRCVATNSA 63
Search completed: March 4, 2003, 16:53:17
Job time : 50.0508 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:51:40 ; Search time 13.7288 Seconds
(without alignments)
135.950 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 229

Sequence: 1 QPLRRQVVVPLSPRLVLL.....RLPLKRMGGSYRCVNASTAN 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	214.5	93.7	1162	1 TCNA_TRYCR	P23253 trypanosoma
2	58.5	25.5	216	1 YLH3_CABEL	P34357 caenorhabdi
3	55	24.0	608	1 HYCC_ECOLI	P16429 escherichia
4	53	23.1	298	1 Y4MQ_RHISN	P55576 rhizobium s
5	51	22.3	289	1 ISPE_RALSO	O8Y2E0 ralstonia s
6	50.5	22.1	177	1 ILIX_RABIT	P26890 oryctolagus
7	50.5	22.1	234	1 YX04_MYCTU	Q50721 mycobacteri
8	50.5	22.1	302	1 CASB_MACEU	P28550 macropus eu
9	49.5	21.6	268	1 YGDL_ECOLI	O46927 escherichia
10	49.5	21.6	305	1 YEDI_ECOLI	P46125 escherichia
11	49	21.4	169	1 RECA_STRMU	P27624 streptococc
12	49	21.4	470	1 SNXH_HUMAN	O15036 homo sapien
13	49	21.4	3767	1 MUAX_CABEL	P34576 caenorhabdi
14	48	21.0	346	1 GP41_HUMAN	O14843 homo sapien
15	48	21.0	346	1 DP41_HUMAN	O15529 homo sapien
16	48	21.0	349	1 DP04_PSEAE	O15534 pseudomonas
17	48	21.0	357	1 DP04_RALSO	O8XZ19 ralstonia s
18	48	21.0	359	1 KMOS_XENLA	P12965 xenopus lae
19	48	21.0	460	1 DNAL_CHLPN	Q9Z8M9 chlamydia p
20	48	21.0	604	1 SP20_YEAST	P50875 saccharomyc
21	48	21.0	610	1 VE1_HPV60	O80943 human papil
22	48	21.0	820	1 MDL2_YEAST	P33311 saccharomyc
23	47.5	20.7	145	1 OAZ_ONCVO	O9NH25 onchocerca
24	47.5	20.7	392	1 NIRE_PSEAE	O51480 pseudomonas
25	47	20.5	169	1 YS4L_PNECA	Q01688 pneumocysti
26	47	20.5	312	1 PYRB_AERPE	O9YB04 aeropyrum p
27	47	20.5	618	1 ORC2_DROME	O24168 drosophila
28	47	20.5	1333	1 ADO_MOUSE	O54754 mus musculu
29	46.5	20.3	778	1 EXP7_STRPN	P35597 streptococc
30	46.5	20.3	894	1 FOX2_NEUCR	Q01373 neurospora
31	46	20.1	135	1 IL4_SHEEP	P30368 ovis aries
32	46	20.1	178	1 YGY5_YEAST	P53071 saccharomyc
33	46	20.1	227	1 Y323_MYCGE	P47565 mycoplasma

RESULT 1

ID	TCNA_TRYCR	STANDARD;	PRT;	1162 AA.
AC	P23253;			
DT	01-NOV-1991	(Rel. 20, Created)		
DT	01-NOV-1991	(Rel. 20, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen).			
GN	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Silvio X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,			
RA	Prioli R.P.;			
RT	"The trypanosoma cruzi neuraminidase contains sequences similar to			
RT	bacterial neuraminidases, YWTD repeats of the low density lipoprotein			
RT	receptor, and type III modules of fibronectin."			
RL	J. Exp. Med. 174:179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION.			
RX	MEDLINE=91376547; PubMed=1896773;			
RA	Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;			
RT	"Trypanosoma cruzi: Localization of neuraminidase on the surface of			
RT	trypanastigotes.";			
RL	Trop. Med. Parasitol. 42:146-150(1991).			
CC	-!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN			
CC	PARASITE INVASION OF CELLS.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-, alpha-(2-8)-glycosidic linkages of terminal sialic residues in			
CC	oligosaccharides, glycoproteins, glycolipids, colominic acid and			
CC	synthetic substrates.			
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR			
CC	(POSSIBLE).			
CC	-!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM			
CC	IN EPIMASTIGOTES, AND NO DETECTABLE IN AMASTIGOTES.			
CC	-!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT			
CC	DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA			
CC	PROTEIN.			
CC	-!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.			
CC	-!- SIMILARITY: CONTAINS 3 BNR REPEATS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; M61732; AAA30255.1; -			
DR	PIR; JH0557; JH0557.			
DR	InterPro; IPR002860; GH_BNR.			

ALIGNMENTS

34	46	20.1	248	1	DATI_YEAST
35	46	20.1	277	1	DEFM_LYCES
36	46	20.1	322	1	HEM2_ARCFU
37	46	20.1	403	1	DP04_CAUCR
38	46	20.1	436	1	BRNO_HAEIN
39	46	20.1	449	1	TRB2_AERPE
40	46	20.1	491	1	AMPG_ECOLI
41	46	20.1	547	1	CITA_KLEPN
42	46	20.1	838	1	AXN2_RAT
43	46	20.1	1333	1	ADO_RAT
44	45.5	19.9	221	1	RL1_SULAC
45	45.5	19.9	2556	1	NTC1_HUMAN

P13483	saccharomyc
O9FU20	lycopersico
O28305	archaeoglob
O9A511	caulobacter
P71345	haemophilus
O9Y9H2	aeropyrum p
P36670	escherichia
P52687	klebsiella
O70240	rattus norv
O9Z0U5	rattus norv
P35024	sulfolobus
P46531	homo sapien

DR Pfam; PF02012; BNR; 2.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 588 FIBROECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 93.78; Score 214.5; DB 1; Length 1162;
Best Local Similarity 97.88; Pred. No. 6.3e-22;
Matches 44; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QPLRRQVVVPLSPRLVLLAFRCQRLP-KRMGGSYRCVNSTAN 45
DB 356 QPLRRQVVVPLSPRLVLLAFRCQRLP-KRMGGSYRCVNSTAN 399

RESULT 2
YLH3 CAEEL STANDARD; PRT; 216 AA.
AC P34357;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C48B4.3 in chromosome III.
GN C48B4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lacroix P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z29117; CAA82374.1; -;
DR PIR; S40720; S40720.
DR WormPep; C48B4.3; CE00490.
KW Hypothetical protein.
SQ SEQUENCE 216 AA; 25567 MW; 245BA9464FDBB75D CRC64;

Query Match 25.5%; Score 58.5; DB 1; Length 216;
Best Local Similarity 35.0%; Pred. No. 0.54;
Matches 14; Conservative 10; Mismatches 11; Indels 5; Gaps 2;

QY 6 QRVVVPLSPRLVLLAFRCQRLP-KRMGGSYRCVNA 41
DB 99 KRISKIPAKERVLLLEQIRQKMGQQPIREFGG-YQVIRA 137

RESULT 3
HYCC ECOLI STANDARD; PRT; 608 AA.
ID HYCC ECOLI
AC P16429; Q46882;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formate hydrogenlyase subunit 3 (FHL subunit 3) (Hydrogenase-3
DE component C).
DE HYCC OR HEVC OR B2723.
GN Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=90251163; PubMed=2187144;
RA Boehm R., Sauter M., Boeck A.;
RT "Nucleotide sequence and expression of an operon in Escherichia coli
RT coding for formate hydrogenlyase components.";
RL Mol. Microbiol. 4:231-243(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC -1- PATHWAY: HYDROGEN METABOLISM; FHL PATHWAY.
CC -1- SUBUNIT: FHL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE
CC FROM FORMATE ARE RELEASED.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17506; CAA35548.1; -;
DR EMBL; U29579; AAA69233.1; -;
DR EMBL; AE000356; AAC75765.1; -;
DR PIR; S08621; S08621.
DR EcoGene; EG10476; hycc.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
KW Oxidoreductase; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 10 25 POTENTIAL.
FT TRANSMEM 44 67 POTENTIAL.
FT TRANSMEM 76 93 POTENTIAL.
FT TRANSMEM 116 140 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 197 218 POTENTIAL.
FT TRANSMEM 229 251 POTENTIAL.
FT TRANSMEM 258 280 POTENTIAL.
FT TRANSMEM 296 312 POTENTIAL.
FT TRANSMEM 416 440 POTENTIAL.
FT TRANSMEM 453 476 POTENTIAL.

DE 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK)

```

RA Brewer M.T., Eisenberg S.P., Ng R.K.;
RT "Rabbit interleukin-1 receptor antagonist. Cloning, expression,
RT functional characterization, and regulation during intestinal
RT inflammation.";
RT J. Biol. Chem. 269:6962-6971(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Hamada H., Mulligan R.C.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052512; PubMed=1427977;
RA Goto F., Goto K., Miyata T., Ohkawa S., Takao T., Mori S.,
RA Furukawa S., Maeda T., Iwanaga S., Shimoinishi Y., Yoshinaga M.;
RT "Interleukin-1 receptor antagonist in inflammatory exudate cells of
RT rabbits. Production, purification and determination of primary
RT structure.";
RL Immunology 77:235-244(1992).
CC -!- FUNCTION: IL-1RA INHIBITS THE ACTIVITY OF IL-1 BY BINDING TO ITS
CC RECEPTOR. IL-1RA HAS NO IL-1 LIKE ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE IL-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S68977; AAB30093.1; -
DR EMBL; M57526; AAA31374.1; -
DR EMBL; D21832; BAA04860.1; -
DR PIR; A54377; A54377.
DR HSP; P18510; IL1R.
DR InterPro; IPR000975; Interleukin_1.
DR Pfam; PF00340; IL1; 1.
DR ProDom; PD002536; Interleukin_1; 1.
DR SMART; SM00125; IL1; 1.
DR PROSITE; PS00253; INTERLEUKIN_1; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 177
FT PROTEIN.
FT INTERLEUKIN-1 RECEPTOR ANTAGONIST
FT BY SIMILARITY.
FT DISULFID 91 141
FT CARBOHYD 109 109
FT N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 177 AA; 20214 MW; F5BC087F097FEAF CRC64;

Query Match 22.1%; Score 50.5; DB 1; Length 177;
Best Local Similarity 36.1%; Pred. No. 5.6;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVVPLSPRLVLLAFRCRQLPLKRMGGSYRCVNA 41
Db BRIDVPLEPQLFLGIQKGLCLS-----CVKS 97

RESULT 7
YY04 MYCTU STANDARD; PRT; 234 AA.
AC Q50721;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV3404c precursor.
GN RV3404C OR MT3512 OR MTCY78.24.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA Collet C., Joseph R., Nicholas K.R.;
RT "Molecular characterization and in-vitro hormonal requirements for

```

```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, TO METHIONYL-TRNA FORMYLTRANSFERASE
CC (EC 2.1.2.9).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77165; CAB01019.1; -
DR EMBL; AE007157; AAK47850.1; -
DR TIGR; MT3512; -
DR TubercuList; RV3404c; -
DR InterPro; IPR002376; formyl_transf.
DR Pfam; PF00551; formyl_transf; 1.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 234
FT SIGNAL 17 234
FT CHAIN 17 234
SQ SEQUENCE 234 AA; 26515 MW; 63FF857BB6FFAA8F CRC64;

Query Match 22.1%; Score 50.5; DB 1; Length 234;
Best Local Similarity 40.0%; Pred. No. 7.5;
Matches 14; Conservative 4; Mismatches 16; Indels 1; Gaps 1;

QY 6 QRVVVVPLSPRLVLLAFRCRQLPLKRMGGSYRCVN 40
Db ERVAEIVERYDLVLSFCKORFFAALIDG-VRCVN 79

RESULT 8
CASB MACEU STANDARD; PRT; 302 AA.
AC P28550;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta casein precursor.
GN CSN2.
OS Macropus eugenii (Tamar wallaby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Macropodidae; Macropus.
OX NCBI_TaxID=9315;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=92181564; PubMed=1543530;
RA Collet C., Joseph R., Nicholas K.R.;
RT "Molecular characterization and in-vitro hormonal requirements for

```



```
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 40.1-50.0 min region on the linkage map.";
RN DNA Res. 3:379-392(1996).
RP [3]
RC SEQUENCE OF 1-136 FROM N.A.
RX STRAIN=K12;
RA MEDLINE=89345179; PubMed=2527357;
RA Hancok T., Gerwin N., Fritz H.-J.;
RT "Nucleotide sequence of the dcm locus of Escherichia coli K12.";
RL Nucleic Acids Res. 17:5844-5844 (1989).
RN [4]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RL three gene classes.";
RN Nucleic Acids Res. 23:3554-3562(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000287; AAC75024.1; -.
DR EMBL; D90835; BAA15785.1; -.
DR EMBL; X13330; -. NOT_ANNOTATED_CDS.
DR EcoGene; EG12709; yedI.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 232 252 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
SQ SEQUENCE 305 AA; 32190 MW; 924454A053A8CEB3 CRC64;

Query Match 21.6%; Score 49.5; DB 1; Length 305;
Best Local Similarity 36.4%; Pred. No. 14;
Matches 12; Conservative 10; Mismatches 8; Indels 3; Gaps 2;

QY 7 RVVVVPLSLRVLLAFRCQRL-PLKMGGSYRC 38
DB 71 KVLVPLA--LIISAFIPWATPLLMIGGAFLC 101

RESULT 11
RECA STRMU
ID RECA STRMU STANDARD; PRT; 169 AA.
AC P27624; Q54467;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RecA protein (Recombinase A) (Fragments).
GN RECA.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE OF 1-112 FROM N.A.
RX MEDLINE=92210521; PubMed=1556091;
RA Dybvig K., Hollingshead S.K., Heath D.G., Clewell D.B., Sun F.,
RA Woodard A.;
RT "Degenerate oligonucleotide primers for enzymatic amplification of
RL recA sequences from Gram-positive bacteria and mycoplasmas.";
RN J. Bacteriol. 174:2729-2732(1992).
RP [2]
RP SEQUENCE OF 113-169 FROM N.A.

RX MEDLINE=923331947; PubMed=1628842;
RA Quivey R.G. Jr., Faustoferri R.C.;
RT "In vivo inactivation of the Streptococcus mutans recA gene mediated
RL by PCR amplification and cloning of a recA DNA fragment.";
RN Gene 116:35-42(1992).
CC -!- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RECA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81468; AAA26929.1; ALT SEQ.
DR EMBL; M61897; -. NOT_ANNOTATED_CDS.
DR HSSP; P26345; IG19.
DR InterPro; IPR001553; RecA.
DR PRINTS; PR00142; RECA.
DR ProDom; PD000229; RecA; 1.
DR PROSITE; PS00321; RECA_1; PARTIAL.
DR PROSITE; PS0162; RECA_2; 1.
DR PROSITE; PS0163; RECA_3; PARTIAL.
KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
FT NON_TER 1 1
FT NON_CONS 102 103
FT NON_TER 169 169
FT NON_TER 169 169
SQ SEQUENCE 169 AA; 18471 MW; 281CF29BA77BARCA CRC64;

Query Match 21.4%; Score 49; DB 1; Length 169;
Best Local Similarity 40.0%; Pred. No. 8.7;
Matches 14; Conservative 5; Mismatches 2; Indels 14; Gaps 2;

QY 3 LRRQRVVVPLS-----PRLVLLA-----FC 23
DB 118 LRRKMAVLPLSQNNPLIQPMLLLALILMSFC 152

RESULT 12
SNXH HUMAN
ID SNXH HUMAN STANDARD; PRT; 470 AA.
AC Q15036;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sorting nexin 17.
GN SNX17 OR KIAA0064.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawaiabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RL analysis of cDNA clones from human cell line KG-1.";
RN DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Wightman P.J., Bonthron D.T.;
RT "Genomic structure of the KIAA064 gene.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
```

```

RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph, Muscle, and Placenta;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP INTERACTION WITH P-SELECTIN.
RX PubMed=11237770;
RT Florian V., Schluter T., Bohnensack R.;
RA "A new member of the sorting nexin family interacts with the
RT C-terminus of P-selectin.";
RL Biochem. Biophys. Res. Commun. 281:1045-1050(2001).
CC -!- FUNCTION: May be involved in several stages of intracellular
CC trafficking (by similarity).
CC -!- SUBUNIT: Interacts with the C-terminus of P-selectin.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31764; BAA06542.1; -
DR EMBL; AJ404855; CAC12897.1; -
DR EMBL; AJ404856; CAC12897.1; JOINED.
DR EMBL; BC002524; AAH02524.1; -
DR EMBL; BC002610; AAH02610.1; -
DR EMBL; BC014620; AAH14620.1; -
DR Genew; HGNC:14979; SNX17.
DR MIM; 605963; -
DR InterPro; IPR001683; PX.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00787; PX; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS0195; PX; 1.
KW Transport; Protein transport.
FT DOMAIN 1 109
FT SEQUENCE 470 AA; 52901 MW; B1C8650CB8AFB5EE CRC64;
SQ
Query Match 21.4%; Score 49; DB 1; Length 470;
Best Local Similarity 34.2%; Pred. No. 25;
Matches 13; Conservative 8; Mismatches 11; Indels 6; Gaps 2;
OY 4 RRRVVVVPLSPRLVLLAFCRQ----RLPLKRMGGSGYR 37
DB 358 RLQWVTIT--SPQAIMMSICLSQSMVDLMVKSGGSIR 393
RESULT 13
MUA3_CAEL STANDARD; PRT; 3767 AA.
ID AC P34576; Q9UAI3; Q21340;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transmembrane cell adhesion receptor mua-3 precursor.
GN MUA-3 OR K08E5.3/T20G5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2.
RT Lu Z., Vogel B., Hedgecock E.;
RA "mu-3 mRNA splicing pattern revealed.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M., Smith A., Kershaw J.K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Involved in cell adhesion.
CC -!- SIMILARITY: CONTAINS 52 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 SEA DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF139060; AAD29428.1; -
DR EMBL; Z30974; CAA83226.2; -
DR EMBL; Z30423; CAA83226.2; JOINED.
DR EMBL; Z30423; CAC42345.1; -
DR EMBL; Z30974; CAC42345.1; JOINED.
DR PIR; S42373; S42373.
DR HSSP; P17301; LAOX.
DR WormPep; K08E5.3; CE28049.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF00008; EGF; 33.
DR Pfam; PF00057; ldl_recept_a; 3.
DR Pfam; PF01390; SEA; 2.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 45.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00200; SEA; 2.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 32.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS50024; SEA; 2.
DR PROSITE; PS50234; VWFA; 2.
KW Receptor; Cell adhesion; EGF-like domain; Repeat; Transmembrane;
FT SIGNAL 1 24
FT CHAIN 25 3767
FT DOMAIN 25 3417
FT DOMAIN 3418 3438
FT DOMAIN 3439 3767
FT DOMAIN 26 63
FT DOMAIN 96 132
FT DOMAIN 133 166
FT DOMAIN 167 209
FT DOMAIN 225 268
FT DOMAIN 375 416
FT DOMAIN 418 466
FT DOMAIN 468 517
FT DOMAIN 519 566
FT DOMAIN 614 663
FT DOMAIN 665 713
FT DOMAIN 714 760
FT DOMAIN 762 810
FT DOMAIN 816 860

```


THIS PAGE BLANK (USPTO)


```

ID Q9HZK3 PRELIMINARY; PRT; 1148 AA.
AC Q9HZK3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcription-repair coupling protein Mfd.
GN MFD OR PA3002.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
CC -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR EMBL; AE004725; AAG06390.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004576; Mfd.
DR InterPro; IPR005118; TRCF.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF03461; TRCF; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00580; mfd; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN_1.
KW ATP-binding; Helicase; Complete Proteome.
SQ SEQUENCE 1148 AA; 128795 MW; DB8BA273C8EB9041 CRC64;

Query Match 26.0%; Score 59.5; DB 16; Length 1148;
Best Local Similarity 33.9%; Pred. No. 4.8;
Matches 19; Conservative 7; Mismatches 13; Indels 17; Gaps 2;

QY 2 PLRRQVVVPLSPRLVLLAFCR-----QRLPLKRM-----GGSYRCVN 40
DB 99 POLKHGLVLPVISTALHRLPTRFLGLGSLVLDVGQKLDVVERMLRLLEGAGYRCVD 154

RESULT 3
O20627 PRELIMINARY; PRT; 94 AA.
ID O20627
AC O20627;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE URFA' protein.
GN URFA'.
OS Physarum polycephalum (slime mold).
OC Mitochondrion.
OC Eukaryota; Mycetozoa; Myxogastria; Myxogastromycetidae; Physarida;
OC Physarum.
OX NCBI_TaxID=5791;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUX2-S;
RX MEDLINE=98177147; PubMed=9508792;
RA Nakagawa C.C., Jones E.P., Miller D.L.;
RT "Mitochondrial DNA rearrangements associated with mF plasmid
RT integration and plasmoidal longevity in Physarum polycephalum.";
RL Curr. Genet. 33:178-187(1998).
```

```

DR EMBL; AF012249; AAC15931.1; -.
KW Mitochondrion.
SQ SEQUENCE 94 AA; 10951 MW; 5326F228977E7531 CRC64;

Query Match 25.8%; Score 59; DB 8; Length 94;
Best Local Similarity 30.9%; Pred. No. 0.44;
Matches 17; Conservative 11; Mismatches 15; Indels 12; Gaps 2;

QY 3 LRQRVVVPLSPRLVLLAF-CRQRLPLKRM-----GGSYRCVNSTAN 45
DB 15 LKRNFLVFPFISTISLLLSWKCRALLRIKLIYSIHVILFIKNTSIRCYSTNSN 69

RESULT 4
Q9FBI1 PRELIMINARY; PRT; 109 AA.
ID Q9FBI1
AC Q9FBI1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Hypothetical 12.4 kDa protein.
OS Shigella dysenteriae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=622;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H2765-39/81;
RX MEDLINE=20407286; PubMed=10948097;
RA Unkmeir A., Schmidt H.;
RT "Structural analysis of phage-borne stx genes and their flanking
RT sequences in shiga toxin-producing escherichia coli and shigella
RT dysenteriae type 1 strains.";
RL Infect. Immun. 68:4856-4864(2000).
DR EMBL; AJ271153; CAC05624.1; -.
KW Hypothetical protein.
SQ SEQUENCE 109 AA; 12388 MW; 136DE5F17722F6A CRC64;

Query Match 25.3%; Score 58; DB 2; Length 109;
Best Local Similarity 35.7%; Pred. No. 0.73;
Matches 15; Conservative 5; Mismatches 14; Indels 8; Gaps 1;

QY 11 VPLSPRLVLLAFCRQRLPLKRMGGSYR-----CVNSTA 44
DB 17 IPCRPLPARAFILSIRKPLRRGAGEWRYVPLHGAGCLPVSSA 58

RESULT 5
Q8Z4B6 PRELIMINARY; PRT; 608 AA.
ID Q8Z4B6
AC Q8Z4B6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Formate hydrogenlyase subunit 3.
GN STY2973.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
```

[illegible]

```

NN SEQUENCE FROM N.A.
RP MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RL cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003605; BAB77501.1; -.
DR InterPro; IPR003439; ABC transporter.
DR PROSITE; PS00211; ABC TRANSPORTER; UNKNOWN 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 677 AA; 76470 MW; 4FA84C2322F4B1BC CRC64;

Query Match 24.7%; Score 56.5; DB 16; Length 677;
Best Local Similarity 36.8%; Pred. No. 7.8;
Matches 14; Conservative 6; Mismatches 5; Indels 13; Gaps 1;

QY 16 RLVL-----LAFCRQLPLKRMGSGYRCVN 40
DB 625 RLVLFYNGYIPWNYARFLDYCTERLFLQVGGYRFIH 662
||||| :|||:|||||
-----LAFCRQLPLKRMGSGYRCVN 40

RESULT 10
Q8XEE2 SEQUENCE FROM N.A.
ID Q8XEE2 PRELIMINARY; PRT; 608 AA.
AC Q8XEE2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Membrane-spanning protein of hydrogenase 3 (part of PHL
DE complex).
GN HYCC OR Z4031 OR ECS3579.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OC Escherichia
OX NCBI_TaxID=83334;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11256796;
RA Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005500; AAG57830.1; -.
DR InterPro; AP002562; BAB37002.1; -.
DR InterPro; IPR003918; NADH ub oxred4.
DR InterPro; IPR001750; Oxidored q1.
DR Pfam; PF00361; oxidored q1; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
KW Complete proteome.
FT CONFLICT 583 S -> G (IN REF. 2).
SQ SEQUENCE 608 AA; 64155 MW; 373E21E813BCEFF8 CRC64;

```

```

Query Match 24.0%; Score 55; DB 16; Length 608;
Best Local Similarity 30.8%; Pred. No. 12;
Matches 12; Conservative 10; Mismatches 15; Indels 2; Gaps 2;

QY 1 QPLRRQVVVVPLSPRLVLLAFCR-QRLPLKRMGSGYRC 38
DB 500 QPMITLLLIACPLLP-FIIMAICKGDRLPSRSGAAWVC 537
||| :|||:|||||
-----QPMITLLLIACPLLP-FIIMAICKGDRLPSRSGAAWVC 537

RESULT 11
Q26966 SEQUENCE FROM N.A.
ID Q26966 PRELIMINARY; PRT; 642 AA.
AC Q26966;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021510; PubMed=7935611;
RA Campetella O.E., Uttaro A.D., Parodi A.J., Frasch A.C.;
RT "A recombinant Trypanosoma cruzi trans-sialidase lacking the amino
RT acid repeats retains the enzymatic activity.";
RL Mol. Biochem. Parasitol. 64:337-340(1994).
DR EMBL; L26499; AAA66382.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70593 MW; 65BD95DD7ADBC222 CRC64;

Query Match 24.0%; Score 55; DB 5; Length 642;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
DB 436 AYRCVNASTAN 446
|||||
-----SYRCVNASTAN 45

RESULT 12
Q26967 SEQUENCE FROM N.A.
ID Q26967 PRELIMINARY; PRT; 642 AA.
AC Q26967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=RA;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RT cruzi trans-sialidases.";
RL Gene 160:123-128(1995).
DR EMBL; L38456; AAA99442.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70453 MW; BC263FAAAF61EA6E CRC64;

Query Match 24.0%; Score 55; DB 5; Length 642;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
DB 436 AYRCVNASTAN 446
|||||
-----SYRCVNASTAN 45

```

RESULT 13

Q26968 PRELIMINARY; PRT; 642 AA.
ID Q26968
AC Q26968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TULAHUEN;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases";
RL Gene 160:123-128(1995).
DR EMBL; L38457; AAA99443.1; -;
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70459 MW; 09F047336492081B CRC64;

Query Match 24.0%; Score 55; DB 5; Length 642;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNSTAN 45

Db 436 AYRCVNSTAN 446

RESULT 14

Q9BHJ5 PRELIMINARY; PRT; 643 AA.
ID Q9BHJ5
AC Q9BHJ5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trans-sialidase.
GN TCTS.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y-STRAIN;
RA Laroy W., Contreras R.;
RT "Cloning of Trypanosoma cruzi trans-sialidase and expression in Pichia pastoris";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ276679; CAC34453.1; -;
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 643 AA; 70546 MW; E46D9A7CE9218ECB CRC64;

Query Match 24.0%; Score 55; DB 5; Length 643;
Best Local Similarity 90.9%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNSTAN 45

Db 437 AYRCVNSTAN 447

RESULT 15

Q8YL12 PRELIMINARY; PRT; 651 AA.
ID Q8YL12
AC Q8YL12;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ALR7126.
GN ALR7126.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003600; BAB78210.1; -;
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 651 AA; 74404 MW; 6AB753997CD2C809 CRC64;

Query Match 24.0%; Score 55; DB 16; Length 651;
Best Local Similarity 42.9%; Pred. No. 13;
Matches 9; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 20 LAFRCRQLPLKRMGSGYRCVN 40

Db 616 LDYCTERMLLQRYGGYRPIH 636

Search completed: March 4, 2003, 16:56:18
Job time : 117.169 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:52:45 ; Search time 18.3051 Seconds
(without alignments)
72.331 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 229

Sequence: 1 QPLRRQVVVVPLSPRLVLL.....RPLKRMGGSYRCVNSTAN 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	24.0	642	4	US-08-911-393-4
2	55	24.0	1060	4	US-08-911-393-2
3	50.5	22.1	177	3	US-09-000-630C-22
4	50.5	22.1	177	3	US-08-862-730C-22
5	50.5	22.1	177	4	US-09-417-455-11
6	50.5	22.1	177	4	US-09-348-942-11
7	50.5	22.1	177	4	US-09-457-626-11
8	48.5	21.2	1436	2	US-08-652-971-2
9	48.5	21.2	1436	2	US-08-991-258A-2
10	48.5	21.2	1436	2	US-08-769-399-2
11	48.5	21.2	1436	3	US-08-991-953A-2
12	48	21.0	401	2	US-08-820-521-2
13	48	21.0	401	4	US-09-248-715-2
14	48	21.0	401	4	US-09-248-715-2
15	47.5	20.7	302	4	US-09-877-730-14
16	47.5	20.7	380	4	US-09-877-730-4
17	47.5	20.7	826	4	US-09-877-730-16
18	47.5	20.7	904	4	US-09-877-730-6
19	47.5	20.7	907	4	US-09-877-730-20
20	47.5	20.7	985	4	US-09-877-730-10
21	47.5	20.7	991	4	US-09-877-730-12
22	47.5	20.7	1069	4	US-09-877-730-2
23	47.5	20.7	1072	4	US-09-877-730-18
24	47.5	20.7	1150	4	US-09-877-730-8
25	47.5	20.7	1638	4	US-09-315-793-12
26	47.5	20.7	3287	2	US-08-477-451-7
27	47	20.5	45	2	US-08-838-957A-24

28	47	20.5	209	4	US-09-134-001C-3980	Sequence 3980, Ap
29	47	20.5	511	4	US-08-991-677-4	Sequence 4, Appli
30	46	20.1	435	4	US-09-446-754-2	Sequence 2, Appli
31	46	20.1	468	4	US-09-446-754-4	Sequence 4, Appli
32	46	20.1	468	4	US-09-446-754-10	Sequence 10, Appli
33	46	20.1	3080	6	5223423-4	Patent No. 5223423
34	45.5	19.9	2556	1	US-08-083-590A-20	Sequence 20, Appl
35	45.5	19.9	2556	3	US-08-532-384-20	Sequence 20, Appl
36	45	19.7	436	3	US-08-486-099-94	Sequence 94, Appl
37	45	19.7	436	3	US-08-360-107A-104	Sequence 104, App
38	45	19.7	436	3	US-08-484-223B-94	Sequence 94, Appl
39	45	19.7	436	3	US-08-319-597-94	Sequence 94, Appl
40	45	19.7	436	3	US-08-475-668A-94	Sequence 94, Appl
41	45	19.7	436	3	US-08-485-551A-94	Sequence 94, Appl
42	45	19.7	436	3	US-08-471-913A-94	Sequence 94, Appl
43	45	19.7	436	4	US-08-485-264A-94	Sequence 94, Appl
44	45	19.7	436	4	US-08-474-349A-94	Sequence 94, Appl
45	45	19.7	436	4	US-08-255-208A-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-08-911-393-4
; Sequence 4, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; SIALYLOLIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; LENGTH: 642 amino acids
; MOLECULE TYPE: protein
US-08-911-393-4

Query Match 24.0%; Score 55; DB 4; Length 642;
Best Local Similarity 90.9%; Pred. No. 3.7;

Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
:|||||
Db 436 AYRCVNASTAN 446

RESULT 2

US-08-911-393-2
; Sequence 2, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPP, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: SIALYLIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-911-393-2

Query Match 24.0%; Score 55; DB 4; Length 1060;
Best Local Similarity 90.9%; Pred. No. 6.8;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
:|||||
Db 467 AYRCVNASTAN 477

RESULT 3

US-09-000-630C-22
; Sequence 22, Application US/09000630C
; Patent No. 6018029
; GENERAL INFORMATION:
; APPLICANT: Fuentes, Gerald M
; APPLICANT: Fuentes, Nelson L.
; TITLE OF INVENTION: DNA Encoding Canine Interleukin-1 Receptor
; TITLE OF INVENTION: Antagonist

Query Match 22.1%; Score 50.5; DB 3; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVPLSRVLLAFRCRORPLKRMGGSYRCVNA 41
:|||||
Db 69 ERIDVVPLEPQLPLGIQKGLCLS-----CVKS 97

RESULT 4

US-08-862-730C-22
; Sequence 22, Application US/08862730C
; Patent No. 6063600
; GENERAL INFORMATION:
; APPLICANT: Fuentes, Gerald M
; APPLICANT: Fuentes, Nelson L.
; TITLE OF INVENTION: DNA Encoding Canine Interleukin-1 Receptor
; TITLE OF INVENTION: Antagonist
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Douglas C Murdock/ Bradley, Arant, Rose & White
; STREET: 2001 Park Place, Suite 1400
; CITY: Birmingham
; STATE: Alabama
; COUNTRY: USA
; ZIP: 35203-2736
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,730C
; FILING DATE: 5/23/97
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: rabbit IL-1ra sequence
; US-08-862-730C-22

Query Match 22.1%; Score 50.5; DB 3; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;

QY 6 QRVVVPLSRVLLAFRCRORPLKRMGGSYRCVNA 41
:|||||
Db 69 ERIDVVPLEPQLPLGIQKGLCLS-----CVKS 97

RESULT 4

US-08-862-730C-22
; Sequence 22, Application US/08862730C
; Patent No. 6063600
; GENERAL INFORMATION:
; APPLICANT: Fuentes, Gerald M
; APPLICANT: Fuentes, Nelson L.
; TITLE OF INVENTION: DNA Encoding Canine Interleukin-1 Receptor
; TITLE OF INVENTION: Antagonist
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Douglas C Murdock/ Bradley, Arant, Rose & White
; STREET: 2001 Park Place, Suite 1400
; CITY: Birmingham
; STATE: Alabama
; COUNTRY: USA
; ZIP: 35203-2736
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch,
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft Windows
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,730C
; FILING DATE: 5/23/97
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: rabbit IL-1ra sequence
; US-08-862-730C-22

Query Match 22.1%; Score 50.5; DB 3; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;

QY 6 QRVVVPLSRVLLAFRCRORPLKRMGGSYRCVNA 41
:|||||
Db 69 ERIDVVPLEPQLPLGIQKGLCLS-----CVKS 97


```
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVPLSPRLVLLAFRCRORLPKRMGGSYRCVNA 41
   : ||||| : : : : :
Db 69 ERIDVPLEPQLFLGIQKGKCLSL-----CVKS 97

RESULT 5
US-09-417-455-11
; Sequence 11, Application US/09417455
; Patent No. 6294655
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Pace, Ann
; TITLE OF INVENTION: A NOVEL INTERLEUKIN-1 RECEPTOR ANTAGONIST AND USES THEREOF
; FILE REFERENCE: 28110/36328
; CURRENT APPLICATION NUMBER: US/09/417,455
; CURRENT FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: US 09/348,942
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: PCT/US99/04291
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/287,210
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/251,370
; PRIOR FILING DATE: 1999-02-17
; PRIOR APPLICATION NUMBER: US 09/229,591
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: US 09/127,698
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: US 09/099,818
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: US 09/082,364
; PRIOR FILING DATE: 1998-05-20
; PRIOR APPLICATION NUMBER: US 09/079,909
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 09/055,010
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-417-455-11

Query Match 22.1%; Score 50.5; DB 4; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVPLSPRLVLLAFRCRORLPKRMGGSYRCVNA 41
   : ||||| : : : : :
Db 69 ERIDVPLEPQLFLGIQKGKCLSL-----CVKS 97

RESULT 6
US-09-348-942-11
; Sequence 11, Application US/09348942
; Patent No. 6337072
; GENERAL INFORMATION:
; APPLICANT: John Ford
; APPLICANT: John Ford
; TITLE OF INVENTION: A NOVEL INTERLEUKIN-1 RECEPTOR ANTAGONIST AND USES THEREOF
; FILE REFERENCE: 28110/35801
; CURRENT APPLICATION NUMBER: US/09/348,942
; CURRENT FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: PCT/US99/04291
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/287,210
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/251,370
; PRIOR FILING DATE: 1999-02-17
; PRIOR APPLICATION NUMBER: US 09/229,591
; PRIOR FILING DATE: 1999-01-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-417-455-11

Query Match 22.1%; Score 50.5; DB 4; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVPLSPRLVLLAFRCRORLPKRMGGSYRCVNA 41
   : ||||| : : : : :
Db 69 ERIDVPLEPQLFLGIQKGKCLSL-----CVKS 97

RESULT 7
US-09-457-626-11
; Sequence 11, Application US/09457626
; Patent No. 6426191
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Pace, Ann
; TITLE OF INVENTION: A NOVEL INTERLEUKIN-1 RECEPTOR ANTAGONIST AND USES THEREOF
; FILE REFERENCE: 28110/36010
; CURRENT APPLICATION NUMBER: US/09/457,626
; CURRENT FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 09/417,455
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: US 09/348,942
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: PCT/US99/04291
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/287,210
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: US 09/251,370
; PRIOR FILING DATE: 1999-02-17
; PRIOR APPLICATION NUMBER: US 09/229,591
; PRIOR FILING DATE: 1999-01-13
; PRIOR APPLICATION NUMBER: US 09/127,698
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: US 09/099,818
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: US 09/082,364
; PRIOR FILING DATE: 1998-05-20
; PRIOR APPLICATION NUMBER: US 09/079,909
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 09/055,010
; PRIOR FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-457-626-11

Query Match 22.1%; Score 50.5; DB 4; Length 177;
Best Local Similarity 36.1%; Pred. No. 3.9;
Matches 13; Conservative 6; Mismatches 10; Indels 7; Gaps 1;

QY 6 QRVVVPLSPRLVLLAFRCRORLPKRMGGSYRCVNA 41
   : ||||| : : : : :
Db 69 ERIDVPLEPQLFLGIQKGKCLSL-----CVKS 97
```

```

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/991,258A
/ FILING DATE: 17-DEC-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/652,971
/ FILING DATE: 24-MAY-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Dreger, Walter H.
/ REGISTRATION NUMBER: 24,190
/ REFERENCE/DOCKET NUMBER: A-63478-3/WH/D/MTK
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1436 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-991-258A-2

Query Match 21.2%; Score 48.5; DB 2; Length 1436;
Best Local Similarity 37.5%; Pred. No. 96;
Matches 18; Conservative 5; Mismatches 10; Indels 15

QY 3 LRQRQVVVP-----LSPRLVLLAFCRQLPLKMGGS----YRCVN 40
      ||| ||| |||
DB 224 LQRQGVLPAGVPHSHRRRLATF-----PLASVGRSEQDLYRCVS 266

RESULT 10
US-08-769-399-2
/ Sequence 2, Application US/08769399
/ Patent No. 5976852
/ GENERAL INFORMATION:
/ APPLICANT: Cheng, Jill
/ APPLICANT: Lasky, Laurence A.
/ TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
/ TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genentech, Inc.
/ STREET: 460 Point San Bruno Blvd.
/ CITY: South San Francisco
/ STATE: California
/ COUNTRY: United States
/ ZIP: 94080
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/769,399
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Dreger, Girger R.
/ REGISTRATION NUMBER: 33,055
/ REFERENCE/DOCKET NUMBER: P1033
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 225-3216
/ TELEFAX: (415) 952-9891
/ TELEX: 910 371-7168
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1436 amino acids

```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-769-399-2

; Query Match          21.2%; Score 48.5; DB 2; Length 1436;
; Best Local Similarity 37.5%; Pred. No. 96;
; Matches 18; Conservative 5; Mismatches 10; Indels 15; Gaps 3;

QY 3 LRQRVVVVP-----LSPLRVLLAFRCRQLPLKRMGGS-----YRCVN 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 LQRQSGVLVPAAGVRHISHRRFLATF-----PLASVGRSEQDLYRCVS 266

RESULT 11
US-08-991-953A-2
; Sequence 2, Application US/08991953A
; Patent No. 6083748
; GENERAL INFORMATION:
; APPLICANT: Cheng, Jill
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT, LLP
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,953A
; FILING DATE: 16-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/652,971
; FILING DATE: 24-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-63478-3/WHD/MTK
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-991-953A-2

; Query Match          21.2%; Score 48.5; DB 3; Length 1436;
; Best Local Similarity 37.5%; Pred. No. 96;
; Matches 18; Conservative 5; Mismatches 10; Indels 15; Gaps 3;

QY 3 LRQRVVVVP-----LSPLRVLLAFRCRQLPLKRMGGS-----YRCVN 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 LQRQSGVLVPAAGVRHISHRRFLATF-----PLASVGRSEQDLYRCVS 266

RESULT 12
US-08-820-521-2
; Sequence 2, Application US/08820521
; Patent No. 5942416
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk
; APPLICANT: Ganesh, Sathe
; APPLICANT: Fuetterer, Wendy
; APPLICANT: Mao, Joyce
; TITLE OF INVENTION: CDNA CLONE HNFYD20 THAT ENCODES
; TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,521
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-820-521-2

; Query Match          21.0%; Score 48; DB 2; Length 401;
; Best Local Similarity 35.1%; Pred. No. 25;
; Matches 13; Conservative 10; Mismatches 12; Indels 2; Gaps 2;

QY 2 PLRRQVVVPLSPRLVLLAFRCRQLP-LKRMGGSYR 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 PVRLEMAVLFVVP-LIITSYCYSRVLVWILGRGGSHR 273

RESULT 13
US-09-248-715-2
; Sequence 2, Application US/09248715
; Patent No. 6207800
; GENERAL INFORMATION:
; APPLICANT: BERGSM, DERK
; APPLICANT: SATHE, GANESH M.
; APPLICANT: FUEETTERER, WENDY
; APPLICANT: MAO, JOYCE
; TITLE OF INVENTION: CDNA CLONE HNFYD20 THAT ENCODES
; TITLE OF INVENTION: A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/248,715
FILING DATE: 09-FEB-1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,521
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-248-715-2

Query Match 21.0%; Score 48; DB 4; Length 401;
Best Local Similarity 35.1%; Pred. No. 25;
Matches 13; Conservative 10; Mismatches 12; Indels 2; Gaps 2;

QY 2 PLRQRVVVPLSPRLVLLAFRCQRLP-LKRMGGSYR 37
DB 238 PVRLEMAVLFVWP-LIITSYCYSLRWILGRGGSHR 273

RESULT 14
US-09-248-715-2
Sequence 2, Application US/09248715
Patent No. 6277960
GENERAL INFORMATION:
APPLICANT: BERGSM, DERK
SATHE, GANESH M.
FUTTERER, WENDY
MAO, JOYCE
TITLE OF INVENTION: CDNA CLONE HNFY20 THAT ENCODES
A NOVEL HUMAN 7-TRANSMEMBRANE RECEPTOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/248,715
FILING DATE: 09-Feb-1999
CLASSIFICATION: UNKNOWN
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,521
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-248-715-2

Query Match 21.0%; Score 48; DB 4; Length 401;
Best Local Similarity 35.1%; Pred. No. 25;
Matches 13; Conservative 10; Mismatches 12; Indels 2; Gaps 2;

QY 2 PLRQRVVVPLSPRLVLLAFRCQRLP-LKRMGGSYR 37
DB 238 PVRLEMAVLFVWP-LIITSYCYSLRWILGRGGSHR 273

RESULT 15
US-09-877-730-14
Sequence 14, Application US/09877730
Patent No. 6465632
GENERAL INFORMATION:
APPLICANT: Walke, D. Wade
APPLICANT: Scoville, John
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
FILE REFERENCE: LEX-0189-USA
CURRENT APPLICATION NUMBER: US/09/877,730
CURRENT FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: US 60/210,607
PRIOR FILING DATE: 2000-06-09
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 302
TYPE: PRT
ORGANISM: homo sapiens
US-09-877-730-14

Query Match 20.7%; Score 47.5; DB 4; Length 302;
Best Local Similarity 24.0%; Pred. No. 21;
Matches 12; Conservative 8; Mismatches 11; Indels 19; Gaps 1;

QY 15 PRVLVLLAFRCQRLPL-----KRMGGSYRCVNASTAN 45
DB 85 PAVITWFEFNRTTLPTMTDRITALPTGVLIQYDVSDSGNRYRCIAIVAH 134

Search completed: March 4, 2003, 16:57:55
Job time : 54.3051 secs

; TITLE OF INVENTION: Methods of Use Therefor

FILE REFERENCE: 1322.1028-001
 CURRENT APPLICATION NUMBER: US/09/745,008
 CURRENT FILING DATE: 2000-12-20
 PRIOR APPLICATION NUMBER: US 60/172,881
 PRIOR FILING DATE: 1999-12-20
 NUMBER OF SEQ ID NOS: 34
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 15
 LENGTH: 21
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Synthetic peptide
 US-09-745-008-15

Query Match 24.0%; Score 55; DB 10; Length 21;
 Best Local Similarity 90.9%; Pred. No. 0.26;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
 :|||||
 Db 11 AYRCVNASTAN 21

RESULT 7

US-09-955-909-4
 Sequence 4, Application US/09955909
 Patent No. US20020150995A1
 GENERAL INFORMATION:
 APPLICANT: PELLETIER, Marc
 BARKER, William A.
 HAKES, David J.
 ZOPF, David A.

TITLE OF INVENTION: METHODS FOR PRODUCING
 SIALYLLOIGOSACCHARIDES IN A DAIRY SOURCE

NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/955,909
 FILING DATE: 18-Sep-2001
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/911,393
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7188-032-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)7909090
 TELEFAX: (212)8699741
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 642 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-955-909-4

Query Match 24.0%; Score 55; DB 10; Length 642;
 Best Local Similarity 90.9%; Pred. No. 12;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
 :|||||
 Db 436 AYRCVNASTAN 446

RESULT 8

US-09-745-008-2
 Sequence 2, Application US/09745008
 Patent No. US20020137667A1
 GENERAL INFORMATION:

APPLICANT: Chuenkova, Marina
 APPLICANT: Pereira, Miercio A.
 TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
 METHODS OF USE THEREFOR
 FILE REFERENCE: 1322.1028-001
 CURRENT APPLICATION NUMBER: US/09/745,008
 CURRENT FILING DATE: 2000-12-20
 PRIOR APPLICATION NUMBER: US 60/172,881
 PRIOR FILING DATE: 1999-12-20
 NUMBER OF SEQ ID NOS: 34
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 666
 TYPE: PRT
 ORGANISM: Trypanosoma cruzi

US-09-745-008-2

Query Match 24.0%; Score 55; DB 10; Length 666;
 Best Local Similarity 90.9%; Pred. No. 12;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SYRCVNASTAN 45
 :|||||
 Db 468 AYRCVNASTAN 478

RESULT 9

US-09-955-909-2
 Sequence 2, Application US/09955909
 Patent No. US20020150995A1
 GENERAL INFORMATION:

APPLICANT: PELLETIER, Marc
 BARKER, William A.
 HAKES, David J.
 ZOPF, David A.

TITLE OF INVENTION: METHODS FOR PRODUCING
 SIALYLLOIGOSACCHARIDES IN A DAIRY SOURCE

NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PENNIE & EDMONDS LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/955,909
 FILING DATE: 18-Sep-2001
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/911,393
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7188-032-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)7909090
 TELEFAX: (212)8699741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1060 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-955-909-2

Query Match 24.0%; Score 55; DB 10; Length 1060;
 Best Local Similarity 90.9%; Pred. No. 20;
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 35 SYRCVNASTAN 45
 Db 467 AYRCVNASTAN 477

RESULT 10

US-09-754-997A-10
 ; Sequence 10, Application US/09754997A
 ; Patent No. US20020102551A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salbaum, Michael J.
 ; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
 ; TITLE OF INVENTION: Acids and Methods of Use
 ; FILE REFERENCE: P-NI 4552
 ; CURRENT APPLICATION NUMBER: US/09/754,997A
 ; CURRENT FILING DATE: 2001-01-04
 ; PRIOR APPLICATION NUMBER: US 60/174,496
 ; PRIOR FILING DATE: 2000-01-04
 ; PRIOR APPLICATION NUMBER: US 60/205,789
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 10
 ; LENGTH: 64
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-754-997A-10

Query Match 23.1%; Score 53; DB 10; Length 64;
 Best Local Similarity 42.1%; Pred. No. 1.7;
 Matches 16; Conservative 6; Mismatches 14; Indels 2; Gaps 2;

Qy 9 VVPLSPRLVLL-AFCRQLPLKRM-GGSYRCVNASTA 44
 Db 26 VTPPEPRLLITPKWLLQILDVQSDAGSYRCVATNSA 63

RESULT 11

US-09-754-997A-4
 ; Sequence 4, Application US/09754997A
 ; Patent No. US20020102551A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salbaum, Michael J.
 ; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
 ; TITLE OF INVENTION: Acids and Methods of Use
 ; FILE REFERENCE: P-NI 4552
 ; CURRENT APPLICATION NUMBER: US/09/754,997A
 ; CURRENT FILING DATE: 2001-01-04
 ; PRIOR APPLICATION NUMBER: US 60/174,496
 ; PRIOR FILING DATE: 2000-01-04
 ; PRIOR APPLICATION NUMBER: US 60/205,789
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 45

SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 932
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-754-997A-4

Query Match 23.1%; Score 53; DB 10; Length 932;
 Best Local Similarity 42.1%; Pred. No. 33;
 Matches 16; Conservative 6; Mismatches 14; Indels 2; Gaps 2;

Qy 9 VVPLSPRLVLL-AFCRQLPLKRM-GGSYRCVNASTA 44
 Db 159 VTPPEPRLLITPKWLLQILDVQSDAGSYRCVATNSA 196

RESULT 12

US-09-908-193-21
 ; Sequence 21, Application US/09908193
 ; Publication No. US20020192748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RASTELLI, LUCA
 ; APPLICANT: SHIMKETS, RICHARD A.
 ; APPLICANT: ZERHUSEN, BRYAN
 ; APPLICANT: MALYANKAR, URIEL M.
 ; APPLICANT: PADIGARU, MURALIDHARA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
 ; FILE REFERENCE: 21402-062
 ; CURRENT APPLICATION NUMBER: US/09/908,193
 ; CURRENT FILING DATE: 2001-07-18
 ; PRIOR APPLICATION NUMBER: 60/220,273
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: 60/221,650
 ; PRIOR FILING DATE: 2000-07-28
 ; PRIOR APPLICATION NUMBER: 60/221,233
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR APPLICATION NUMBER: 60/220,912
 ; PRIOR FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: 60/218,875
 ; PRIOR FILING DATE: 2000-07-18
 ; PRIOR APPLICATION NUMBER: 60/218,870
 ; PRIOR FILING DATE: 2000-07-18
 ; PRIOR APPLICATION NUMBER: 60/218,901
 ; PRIOR FILING DATE: 2000-07-18
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 1252
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-908-193-21

Query Match 23.1%; Score 53; DB 9; Length 1252;
 Best Local Similarity 42.1%; Pred. No. 46;
 Matches 16; Conservative 6; Mismatches 14; Indels 2; Gaps 2;

Qy 9 VVPLSPRLVLL-AFCRQLPLKRM-GGSYRCVNASTA 44
 Db 180 VTPPEPRLLITPKWLLQILDVQSDAGSYRCVATNSA 217

RESULT 13

US-09-754-997A-2
 ; Sequence 2, Application US/09754997A
 ; Patent No. US20020102551A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salbaum, Michael J.
 ; TITLE OF INVENTION: NOPE Polypeptides, Encoding Nucleic
 ; TITLE OF INVENTION: Acids and Methods of Use
 ; FILE REFERENCE: P-NI 4552
 ; CURRENT APPLICATION NUMBER: US/09/754,997A
 ; CURRENT FILING DATE: 2001-01-04
 ; PRIOR APPLICATION NUMBER: US 60/174,496

; PRIOR FILING DATE: 2000-01-04
; PRIOR APPLICATION NUMBER: US 60/205,789
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1252
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-754-997A-2

Query Match 23.1%; Score 53; DB 10; Length 1252;
Best Local Similarity 42.1%; Pred. No. 46;
Matches 16; Conservative 6; Mismatches 14; Indels 2; Gaps 2;

QY 9 VVVPLSPRLVLL-AFCRQLPLKRM-GGSYRCVNSTA 44
Db 180 VTPEERLITLPKWLQILDVQSDAGSYRCVATNSA 217

RESULT 14
US-09-738-626-4710
; Sequence 4710, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4710
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4710

Query Match 22.9%; Score 52.5; DB 9; Length 318;
Best Local Similarity 37.5%; Pred. No. 12;
Matches 12; Conservative 9; Mismatches 8; Indels 3; Gaps 2;

QY 8 VVVPLSPRLVLLAFCRQL-PLKRMGGSYRC 38
Db 77 IILPIA--LLLSAFAPWALTPILMLGGSYLC 106

RESULT 15
US-10-231-353-14
; Sequence 14, Application US/10231353
; Publication No. US20030023064A1
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuln, Alejandro
; APPLICANT: Zambrowicz, Brian

; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20030023064A1el Human Phosphatases and Polynucleotides Enco
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/10/231,353
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US/09/877,730
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 302
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-231-353-14

Query Match 20.7%; Score 47.5; DB 9; Length 302;
Best Local Similarity 24.0%; Pred. No. 54;
Matches 12; Conservative 8; Mismatches 11; Indels 19; Gaps 1;

QY 15 PRLVLLAFCRQLPL-----KRMGGSYRCVNSTAN 45
Db 85 PAVITWEFNRTTLPMTMDRITALPTGVLQIYDVSRDSDGNRYCIAATVAH 134

Search completed: March 4, 2003, 16:57:58
Job time : 27.7797 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:53:45 ; Search time 101.441 Seconds
(without alignments)
59.111 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 45
Sequence: 1 QPLRRQVVVPLSPRLVLL.....RLPLKRMGGSYRCVNSTAN 45

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	22.2	642	20	AA1980
2	10	22.2	669	23	AA1981
3	10	22.2	1060	20	AA1982
4	8	17.8	28	22	AA1983
5	7	15.6	70	22	AA1984
6	7	15.6	231	22	AA1985
7	7	15.6	386	21	AA1986
8	7	15.6	410	22	AA1987
9	7	15.6	438	18	AA1988
10	7	15.6	443	21	AA1989

11	7	15.6	452	18	AA15091	Human precursor pr
12	7	15.6	480	22	AA176860	Human lung tumour
13	7	15.6	480	23	AA185515	Clone #18991 of lu
14	7	15.6	492	21	AA158436	Lung cancer associ
15	7	15.6	577	22	AA136167	Klebsiella pneumon
16	6	13.3	14	22	AA197946	Human peptide #122
17	6	13.3	21	21	AA180349	Bovine factor V th
18	6	13.3	36	22	AA161434	Human TANGO 275 EG
19	6	13.3	41	20	AA188946	Sequence ID #646 f
20	6	13.3	41	22	AA151017	Human secreted pro
21	6	13.3	43	22	AA139504	Peptide #7010 enco
22	6	13.3	43	22	AA124254	Protein #6253 enco
23	6	13.3	43	22	AA160196	Human brain expres
24	6	13.3	43	22	AA172816	Human bone marrow
25	6	13.3	43	22	AA119749	Peptide #6183 enco
26	6	13.3	43	22	AA133046	Peptide #7083 enco
27	6	13.3	43	23	AA142644	Human peptide enco
28	6	13.3	45	21	AA125733	Human secreted pro
29	6	13.3	49	20	AA119530	Amino acid sequenc
30	6	13.3	51	22	AA126230	Novel human diagno
31	6	13.3	52	23	AA103721	Human ORFX protein
32	6	13.3	54	22	AA130444	Peptide #3095 enco
33	6	13.3	54	22	AA135612	Peptide #3118 enco
34	6	13.3	54	22	AA121041	Protein #3040 enco
35	6	13.3	54	22	AA156429	Human brain expres
36	6	13.3	54	22	AA168810	Human bone marrow
37	6	13.3	54	22	AA116628	Peptide #3062 enco
38	6	13.3	54	22	AA129114	Peptide #3151 enco
39	6	13.3	54	22	AA104345	Peptide #3027 enco
40	6	13.3	54	23	AA138389	Human peptide enco
41	6	13.3	55	22	AA143467	Peptide #10973 enc
42	6	13.3	55	22	AA126433	Protein #8432 enco
43	6	13.3	55	22	AA164397	Human brain expres
44	6	13.3	55	22	AA177217	Human bone marrow
45	6	13.3	55	22	AA121150	Peptide #7584 enco

ALIGNMENTS

RESULT 1
AA101541
ID AA101541 standard; Protein; 642 AA.

XX AA101541;

XX 15-JUN-1999 (first entry)

XX Alpha(2-3) trans-sialidase sequence lacking amino acid repeats.

XX Alpha(2-3)trans-sialidase; sialyl-oligosaccharide; sialyllactose;
cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
duodenal ulcer; arthritis; enterotoxin.

XX Trypanosoma cruzi.

XX WO9908511-A1.

XX 25-FEB-1999.

XX 13-AUG-1998; 98WO-US16756.

XX 14-AUG-1997; 97US-0911393.

XX (NEOS-) NEOSE TECHNOLOGIES INC.

XX Barker WA, Hakes DJ, Pelletier M, Zopf DA;

XX WPI; 1999-190079/16.

XX N-PSDB; AA126612.

XX Production of sialyl-oligosaccharides, particularly sialyl-lactose
PT - by treating a dairy source such as a cheese processing waste

stream with an alpha (2-3) trans-sialidase

Disclosure; Fig 4; 84pp; English.

The present sequence represents a functional Trypanosoma cruzi alpha(2-3)trans-sialidase which lacks amino acid repeats. The protein is used in the method of the invention to produce sialyl-oligosaccharides, particularly sialyllactose, which are produced by treating a dairy source such as a cheese processing waste strain with an alpha (2-3) trans-sialidase. The method can be used for producing trans-sialidase. The method can be used for producing sialyl-oligosaccharides, such as (2-3)sialyllactose for pharmaceutical use. (2-3)sialyllactose has been shown to neutralise enterotoxins of various pathogenic microbes including E. coli, Vibrio cholerae and Salmonella. It has also been shown that alpha(2-3)(2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc) interferes with colonisation of Helicobacter pylori and thereby prevents or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has additionally been proposed to inhibit immune complex formation by disrupting occupancy of the Fc carbohydrate binding site on IgG and to be useful in treating arthritis.

Query Match 22.2%; Score 10; DB 20; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNSTAN 45
|||||
Db 437 YRCVNSTAN 446

RESULT 2
ID ABB08420 standard; Protein; 669 AA.
AC ABB08420;
XX
DT 01-JUL-2002 (first entry)
XX
DE Catalytic trans-sialidase unit of T. cruzi amino acid sequence.
XX
KW Mycoplasma associated disease; cell proliferation; trans-sialidase;
KW enzyme; atherosclerotic vascular disease; malignancy;
KW sialic acid; antiatherosclerotic; antibacterial; antiviral; anti-HIV;
KW cytosolic; vasotropic; ovarian carcinoma; breast cancer;
KW prostate cancer; colon cancer; lung cancer; leukaemia; HIV;
KW human immunodeficiency virus; chlamydia; PCR primer.
XX
OS Trypanosoma cruzi.
OS Synthetic.
XX
PN WO200202050-A2.
XX
PD 10-JAN-2002.
XX
PF 03-JUL-2001; 2001WO-BR000083.
XX
PR 03-JUL-2000; 2000BR-0002989.
XX
PA (HIGU/) HIGUCHI M D L.
PA (SCHE/) SCHENKMAN S.
XX
PI Higuchi MDL, Schenkman S;
XX
DR WPI; 2002-154675/20.
DR N-PSDB; ABA98876.
XX
PT Composition useful for treatment of mycoplasma infection and diseases
PT associated with cell proliferation e.g. malignancy or with co-infection
PT with another microbe, comprises agent inhibiting sialic acid-mediated
PT attachment of mycoplasma
XX

Claim 6; Fig 26; 63pp; English.

The invention relates to a composition useful for treating or preventing mycoplasma infection in a subject suffering from a disorder characterised by increased cell proliferation or by co-infection with a second microbe, comprising an agent that prevents or inhibits sialic acid-mediated attachment of mycoplasma to the subject's cells. The activity of compositions of the invention may be described as; antiatherosclerotic, antibacterial, antiviral, anti-HIV, cytostatic and vasotropic. The compositions are useful to treat diseases associated with undesirable cell proliferation, such as atherosclerotic vascular disease and malignancies, by reducing or preventing mycoplasma infection. Examples of malignancies include; ovarian carcinoma, breast cancer, prostate cancer, colon cancer, lung cancer and leukaemia. They are also useful to treat diseases associated with infection with other infectious organisms co-occurring with mycoplasma (and typically increasing the virulence of both pathogens), especially human immunodeficiency virus or chlamydia species. They can be used to treat such diseases in humans or other animals, and can be administered in conjunction with conventional agents e.g. anti-platelet or chemotherapeutic agents. The current sequence represents the catalytic trans-sialidase unit of T. cruzi amino acid sequence.

Query Match 22.2%; Score 10; DB 23; Length 669;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNSTAN 45
|||||
Db 456 YRCVNSTAN 465

RESULT 3
ID AAY01540 standard; Protein; 1060 AA.
XX
AC AAY01540;
XX
DT 15-JUN-1999 (first entry)
XX
DE Trypanosoma cruzi alpha(2-3) trans-sialidase amino acid sequence.
XX
KW Alpha(2-3)trans-sialidase; sialyl-oligosaccharide; sialyllactose;
KW cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW duodenal ulcer; arthritis; enterotoxin.
XX
OS Trypanosoma cruzi.
XX
PN WO9908511-A1.
XX
PD 25-FEB-1999.
XX
PF 13-AUG-1998; 98WO-US16756.
XX
PR 14-AUG-1997; 97US-0911393.
XX
PA (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI Barker WA, Hakes DJ, Pelletier M, Zopf DA;
XX
DR WPI; 1999-190079/16.
DR N-PSDB; AAX26611.
XX
PT Production of sialyl-oligosaccharides, particularly sialyllactose
PT - by treating a dairy source such as a cheese processing waste
PT stream with an alpha (2-3) trans-sialidase
XX
XX Disclosure; Fig 2; 84pp; English.
XX
PT The present sequence represents Trypanosoma cruzi alpha(2-3)trans-
CC sialidase. The protein is used in the method of the invention to

CC produce sialyl-oligosaccharides, particularly sialyllactose, which
 CC are produced by treating a dairy source such as a cheese processing
 CC waste strain with an alpha (2-3) trans-sialidase. The method can be
 CC used for producing sialyl-oligosaccharides, such as (2-3)sialyllactose
 CC for pharmaceutical use. (2-3)sialyllactose has been shown to neutralise
 CC enterotoxins of various pathogenic microbes including E. coli, Vibrio
 CC cholerae and Salmonella. It has also been shown that
 CC alpha(2-3)(2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc)
 CC interferes with colonisation of Helicobacter pylori and thereby prevents
 CC or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has
 CC additionally been proposed to inhibit immune complex formation by
 CC disrupting occupancy of the Fc carbohydrate binding site on IgG and to
 CC be useful in treating arthritis.

XX
 SQ Sequence 1060 AA;

Query Match 22.2%; Score 10; DB 20; Length 1060;
 Best Local Similarity 100.0%; Pred. No. 0.2; Mismatches 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0;

QY 36 YRCVNSTAN 45
 |||||
 Db 468 YRCVNSTAN 477

RESULT 4
 AAO09185
 ID AAO09185 standard; Protein; 28 AA.
 XX
 AC AAO09185;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 23077.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.
 XX
 XX WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI89116.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX
 PS Claim 20; SEQ ID NO 23077; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 28 AA;

Query Match 17.8%; Score 8; DB 22; Length 28;
 Best Local Similarity 100.0%; Pred. No. 0.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PLSPRLVL 19
 |||||
 Db 1 PLSPRLVL 8

RESULT 5
 AAO12427
 ID AAO12427 standard; Protein; 70 AA.
 XX
 AC AAO12427;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 26319.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.
 XX
 XX WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI92358.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX
 PS Claim 20; SEQ ID NO 26319; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 70 AA;

Query Match 15.6%; Score 7; DB 22; Length 70;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 VPLSPRL 17
Db 41 VPLSPRL 47
|||||

RESULT 6
ABB70904
ID ABB70904 standard; Protein; 231 AA.
XX
AC ABB70904;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39504.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PA 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL15007.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 39504; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 231 AA;

Query Match 15.6%; Score 7; DB 22; Length 231;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LRRQRW 9
Db 122 LRRQRW 128
|||||

RESULT 7
AAY78589
ID AAY78589 standard; Protein; 386 AA.
XX

AC AAY78589;
XX
DT 05-MAY-2000 (first entry)
XX
DE Metalloprotease MIFR clone protein sequence.
XX
KW Metalloprotease in the female reproductive tract; MIFR; human; MMP;
KW matrix metalloprotease.
XX
OS Homo sapiens.
XX
PN JP2000014387-A.
XX
PD 18-JAN-2000.
XX
PF 06-JUL-1998; 98JP-0190869.
XX
PR 06-JUL-1998; 98JP-0190869.
XX
PA (TAKA/) TAKAHASHI T.
PA (SDIS-) SDI KK.
XX
DR WPI; 2000-154341/14.
DR N-PSDB; AA290014.
XX
PT A new metalloprotease and a DNA coding it -
XX
PS Disclosure; Page 19-20; 21pp; Japanese.
XX
CC This sequence represents a protein sequence of a human metalloprotease in
CC the female reproductive tract (MIFR) clone. MIFR is a matrix
CC metalloprotease (MMP). The invention relates to the metalloprotease
CC protein which is 390 amino acids in length. A recombinant vector
CC containing the MIFR gene can be used to create transformants which
CC produce the metalloprotease in culture.
XX
SQ Sequence 386 AA;

Query Match 15.6%; Score 7; DB 21; Length 386;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PLRRQRV 8
Db 202 PLRRQRV 208
|||||

RESULT 8
ABB70903
ID ABB70903 standard; Protein; 410 AA.
XX
AC ABB70903;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39501.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PA 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX

PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL15006.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 39501; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 410 AA;
 Query Match 15.6%; Score 7; DB 22; Length 410;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LRQRVV 9
 Db 315 LRQRVV 321
 |||||
 RESULT 9
 AAW15092
 ID AAW15092 standard; protein; 438 AA.
 XX
 AC AAW15092;
 XX
 DT 18-DEC-1997 (first entry)
 XX
 DE Human protective protein cathepsin A.
 XX
 KW Protective protein cathepsin A; PPCA; precursor; human;
 KW hanging drop crystallisation; diffusion crystallisation;
 KW x-ray crystallography; diffraction; computer modelling;
 KW three-dimensional structure; lysosomal storage disease;
 KW galactosialidosis.
 XX
 OS Homo sapiens.
 XX
 PN WO9715588-A1.
 XX
 PD 01-MAY-1997.
 XX
 PF 25-OCT-1996; 96WO-US17325.
 XX
 PR 15-NOV-1995; 95US-0006802.
 PR 26-OCT-1995; 95US-0009976.
 XX
 PA (DAZZ/) DAZZO A.
 PA (HOLW/) HOL W G J.
 PA (RUDE/) RUDENKO G.
 XX
 PI D'Azzo A, Hol WGJ, Rudenko G;
 XX WPI; 1997-258950/23.
 DR
 XX Crystallising a human protective protein/cathepsin A or precursor -
 PT to determine three dimensional structure by x-ray crystallography
 PT and computer analysis to identify ligands for PPCA-related
 PT pathologies

XX
 PS Claim 10; Fig 14; 118pp; English.
 XX
 CC A new method has been developed for crystallising a human protective
 CC protein/cathepsin A (PPCA) or precursor (ppPCA). The method involves
 CC hanging drop or diffusion crystallisation of purified PPCA or ppPCA
 CC to produce a biologically active product resolvable by x-ray
 CC crystallography to give x-ray diffraction patterns suitable for
 CC three-dimensional structure determination. The present sequence
 CC represents human PPCA. Identification of PPCA/ppPCA three dimensional
 CC structures is useful to delineate specific biological activities and
 CC ligands for rational drug design (RDD) for diagnosis and therapy of
 CC PPCA-related pathologies. RDD by computer modelling can identify ligands
 CC calculated to associate with, or bind to, PPCA/ppPCA sites or domains,
 CC and the ligands screened for biological activity, synthesised/
 CC recombinantly produced and used to diagnose, treat or prevent PPCA
 CC related pathologies in animals, especially humans. PPCA is the main
 CC genetic defect underlying the lysosomal storage disease
 CC galactosialidosis. PPCA/ppPCA ligands may also modulate PPCAs/ppPCAs
 CC from other eukaryotes. The x-ray diffraction patterns produced are of
 CC sufficiently high resolution to determine PPCA/ppPCA three dimensional
 CC structure not previously published; determination of three-dimensional
 CC structure may allow identification of therapeutic or diagnostic ligands
 CC for different PPCA/ppPCAs if significant sequence and structural element
 CC conservation exists.
 XX
 SQ Sequence 438 AA;
 Query Match 15.6%; Score 7; DB 18; Length 438;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 RLPLKRM 32
 Db 279 RLPLKRM 285
 |||||
 RESULT 10
 AAB32529
 ID AAB32529 standard; Protein; 443 AA.
 XX
 AC AAB32529;
 XX
 DT 19-JAN-2001 (first entry)
 XX
 DE S. lavendulae MmcP encoded protein sequence.
 XX
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide.
 XX
 OS Streptomyces lavendulae.
 XX
 PN WO200053737-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 10-MAR-2000; 2000WO-US06394.
 XX
 PR 12-MAR-1999; 99US-0266965.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (SHER/) SHERMAN D H.
 PA (MAOY/) MAO Y.
 PA (VARO/) VAROGLU M.
 PA (HENM/) HE M.
 PA (SHEL/) SHELTON P C.
 XX
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 XX WPI; 2000-601980/57.
 DR N-PSDB; AAC55830.
 DR

XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitomycin ring system biosynthesis -
XX
PS Disclosure; Page 380-382; 399pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzoquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC spanning 55kb of DNA. The invention includes an expression cassette
CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC and host cells transformed with the cassette. The nucleotide, and protein
CC sequences and the transformed host cells of the invention result in
CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC antibiotic activities. The nucleotide sequences are used to elucidate the
CC molecular basis for the biosynthesis of the mitomycin ring system, as well
CC as to engineer the biosynthesis of novel natural products, e.g.
CC antibiotics, anti-inflammatory agents, anti-cancer agents,
CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC obstructive pulmonary disease as well as other disease involving
CC respiratory inflammation, or cholesterol-lowering agents or as crop
CC protection agents (e.g. fungicides or insecticides) as well as
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
CC PHA monomer synthases. Sequences AAC5782-C55881, AAC55815-C55849 and
CC AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC sequences and encoded proteins. Sequences AAC55812-C55814,
CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC cloning of the mitomycin biosynthetic genes.
XX
SQ Sequence 443 AA;
Query Match 15.6%; Score 7; DB 21; Length 443;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 RQRVVVV 11
DB 4 RQRVVVV 10
RESULT 11
AAW15091
ID AAW15091 standard; protein; 452 AA.
XX
AC AAW15091;
XX
DT 18-DEC-1997 (first entry)
XX
DE Human precursor protective protein cathepsin A.
XX
KW Protective protein cathepsin A; PPCA; precursor; human;
KW hanging drop crystallisation; diffusion crystallisation;
KW x-ray crystallography; diffraction; computer modelling;
KW three-dimensional structure; lysosomal storage disease;
KW galactosialidosis.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Peptide 285..298
FT /label= Excision peptide
FT /note= "For conversion to the mature protective
FT protein cathepsin A"
XX
XX WO9715588-A1.
XX
XX 01-MAY-1997.
XX
XX 25-OCT-1996; 96WO-US17325.
XX

PR 15-NOV-1995; 95US-0006802.
PR 26-OCT-1995; 95US-0005976.
XX
PA (DAZZ/) DAZZO A.
PA (HOLW/) HOL W G J.
PA (RUDE/) RUDENKO G.
XX
PI D'Azzo A, Hol W G J, Rudenko G;
XX
DR WPI; 1997-258950/23.
XX
PT Crystallising a human protective protein/cathepsin A or precursor
PT to determine three dimensional structure by x-ray crystallography
PT and computer analysis to identify ligands for PPCA-related
PT pathologies
XX
PS Claim 10; Fig 13; 118pp; English.
XX
CC A new method has been developed for crystallising a human protective
CC protein/cathepsin A (PPCA) or precursor (pPPCA). The method involves
CC hanging drop or diffusion crystallisation of purified PPCA or pPPCA
CC to produce a biologically active product resolvable by x-ray
CC crystallography to give x-ray diffraction patterns suitable for
CC three-dimensional structure determination. The present sequence
CC represents human pPPCA. Identification of PPCA/ppPCA three dimensional
CC structures is useful to delineate specific biological activities and
CC ligands for rational drug design (RDD) for diagnosis and therapy of
CC PPCA-related pathologies. RDD by computer modelling can identify ligands
CC calculated to associate with, or bind to, PPCA/ppPCA sites or domains,
CC and the ligands screened for biological activity, synthesised/
CC recombinantly produced and used to diagnose, treat or prevent PPCA
CC related pathologies in animals, especially humans. PPCA is the main
CC genetic defect underlying the lysosomal storage disease
CC galactosialidosis. PPCA/ppPCA ligands may also modulate PPCAs/ppPCAs
CC from other eukaryotes. The x-ray diffraction patterns produced are of
CC sufficient high resolution to determine PPCA/ppPCA three dimensional
CC structure not previously published; determination of three-dimensional
CC structure may allow identification of therapeutic or diagnostic ligands
CC for different PPCA/ppPCAs if significant sequence and structural element
CC conservation exists.
XX
SQ Sequence 452 AA;
Query Match 15.6%; Score 7; DB 18; Length 452;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RLPLKRM 32
DB 279 RLPLKRM 285
RESULT 12
AAB76860
ID AAB76860 standard; Protein; 480 AA.
XX
AC AAB76860;
XX
DT 12-APR-2001 (first entry)
XX
DE Human lung tumour protein related protein sequence SEQ ID NO:336.
XX
KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW cytostatic; antisense inhibition.
XX
OS Homo sapiens.
XX
XX WO200100828-A2.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000WO-US18061.


```

XX PR 30-JUN-1999; 99US-0346492.
XX PR 15-OCT-1999; 99US-0419356.
XX PR 17-DEC-1999; 99US-0466867.
XX PR 30-DEC-1999; 99US-0476300.
XX PR 06-MAR-2000; 2000US-0519642.
XX PR 22-MAR-2000; 2000US-0533077.
XX PR 10-APR-2000; 2000US-0546259.
XX PR 27-APR-2000; 2000US-0560406.
XX PR 05-JUN-2000; 2000US-0589184.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX PI Retter MW, Mannion J;
XX DR WPI; 2001-071488/08.
XX PT Lung tumor-associated proteins and the nucleic acids that encode them,
XX PT useful for preventing, diagnosing and treating lung cancer -
XX PS Example 1; Page 267-268; 436pp; English.
XX CC The present invention describes immunogenic portions of lung tumour-
XX CC associated proteins (I) and the nucleic acids (NAS) that encode them.
XX CC (I) have cytostatic activity and can be used in gene therapy, antisense
XX CC inhibition and in vaccines. The NAS and the lung tumour-associated
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with their inappropriate expression,
XX CC especially lung cancers. For example, the NAS may be administered to
XX CC treat diseases by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of the protein by expressing inactive proteins
XX CC or to supplement the patients own production of (I). Additionally, the
XX CC NAS may be used to produce the lung-tumour associated protein, according
XX CC to standard recombinant DNA methodology. Conversely, antisense NA
XX CC molecules may be administered to down regulate protein expression by
XX CC binding with the cells own genes and preventing their expression. The NA
XX CC and complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar NA sequences in
XX CC samples, and hence which patients may be in need of treatment for lung
XX CC cancer. The (I) may be used as antigens in the production of antibodies
XX CC and in assays to identify modulators (agonists and antagonists) of the
XX CC expression and activity of the protein. AAF68083 to AAF68878 and
XX CC AAB76848 to AAB76878 represent human lung tumour protein related
XX CC nucleotide and protein sequences which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 480 AA;
Query Match 15.6%; Score 7; DB 22; Length 480;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RLPLKRM 32
DB 307 RLPLKRM 313
RESULT 13
AAU85515
ID AAU85515 standard; Protein; 480 AA.
XX AC AAU85515;
XX DT 21-MAY-2002 (first entry)
XX DE Clone #18991 of lung tumour protein.
XX KW Lung tumour; cancer; T cell; immune response stimulator;
XX KW cytostatic.
XX OS Homo sapiens.
XX PN WO200055180-A2.

```

```

PN WO200204514-A2.
XX 17-JAN-2002.
XX PF 10-JUL-2001; 2001WO-US22058.
XX PR 11-JUL-2000; 2000US-0614124.
XX PR 29-AUG-2000; 2000US-0651563.
XX PR 08-SEP-2000; 2000US-0658824.
XX PR 26-SEP-2000; 2000US-0671325.
XX PR 06-OCT-2000; 2000US-0677419.
XX PR 30-OCT-2000; 2000US-0702705.
XX PR 13-DEC-2000; 2000US-0736457.
XX PR 03-MAY-2001; 2001US-0849626.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
XX PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
XX PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
XX DR WPI; 2002-164634/21.
XX DR N-PSDB; ABK38015.
XX PT Novel polynucleotide encoding a lung tumour polypeptide useful for
XX PT stimulating and/or expanding T cells specific for a tumour protein -
XX PS Example 1; SEQ ID No 336; 223pp; English.
XX CC The invention describes an isolated polynucleotide and polypeptide
XX CC useful for stimulating and/or expanding T cells specific for a tumour
XX CC protein for determining the presence of a cancer in a patient. A
XX CC composition containing the polynucleotide and/or polypeptide is useful
XX CC for treating a lung cancer in a patient. The polypeptide is useful for
XX CC removing tumour cells from a biological sample. The polynucleotide is
XX CC also useful as probe or primer to detect the level of mRNA encoding a
XX CC tumour protein. This is the amino acid sequence of a lung tumour
XX CC associated protein, described in the method of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 480 AA;
Query Match 15.6%; Score 7; DB 23; Length 480;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RLPLKRM 32
DB 307 RLPLKRM 313
RESULT 14
AAB58436
ID AAB58436 standard; Protein; 492 AA.
XX AC AAB58436;
XX DT 14-MAR-2001 (first entry)
XX DE Lung cancer associated polypeptide sequence SEQ ID 774.
XX KW Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnerary;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease.
XX OS Homo sapiens.
XX PN WO200055180-A2.

```

```
XX 21-SEP-2000.
PD
PP 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
DR WPI; 2000-587514/55.
DR N-PSDB; AAF18312.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 11; Page 1301-1302; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnery; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 492 AA;
Query Match 15.6%; Score 7; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RLPLKRM 32
DB 319 RLPLKRM 325
|||||
RESULTS
AAU36167
ID AAU36167 standard; Protein; 577 AA.
XX
XX AAU36167;
XX
XX 14-FEB-2002 (first entry)
XX
XX Klebsiella pneumoniae cellular proliferation protein #155.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Klebsiella pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR
```

```
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS54026.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 11760; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 577 AA;
Query Match 15.6%; Score 7; DB 22; Length 577;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 RVVWVPL 13
DB 402 RVVWVPL 408
|||||
Search completed: March 4, 2003, 17:00:08
Job time : 103.441 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:58:06 ; Search time 19.8305 Seconds
(without alignments)
218.151 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 45

Sequence: 1 QPLRRQVVVVPLSPRLVLL.....RLPLKRMGGSYRCVNSTAN 45

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000.

Post-processing: Listing first 45 summaries

Database : PIR 73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	62.2	1162	2 JH0557	exo-alpha-sialidas
2	7	15.6	85	2 S09541	gene 1.6 protein -
3	7	15.6	120	2 F83473	hypothetical prote
4	7	15.6	148	2 B70964	hypothetical prote
5	7	15.6	238	2 AE2213	two-component resp
6	7	15.6	266	2 AC3297	ABC transporter At
7	7	15.6	268	2 T30630	hypothetical prote
8	7	15.6	377	2 AC2336	potassium-dependen
9	7	15.6	480	2 A31589	carboxypeptidase C
10	7	15.6	580	2 T44481	aerobactin biosynt
11	7	15.6	580	2 S50883	aerobactin biosynt
12	7	15.6	1078	2 E75407	isoleucyl-tRNA syn
13	6	13.3	24	2 E39690	neural cell adhesi
14	6	13.3	29	2 G39690	neural cell adhesi
15	6	13.3	38	2 H39690	neural cell adhesi
16	6	13.3	43	2 I39690	neural cell adhesi
17	6	13.3	49	2 S25433	neural cell adhesi
18	6	13.3	55	2 AC2303	hypothetical prote
19	6	13.3	69	2 S07516	gene 5.7 protein -
20	6	13.3	69	2 S42314	gene 5.7 protein -
21	6	13.3	69	2 A82710	hypothetical prote
22	6	13.3	91	2 A59002	cryptdin-related p
23	6	13.3	91	2 I61586	cysteine-rich cryp
24	6	13.3	93	2 I49102	cryptdin 3 - mouse
25	6	13.3	93	2 I49103	cryptdin 5 - mouse
26	6	13.3	93	2 I49104	cryptdin 6 - mouse
27	6	13.3	94	2 F69256	hypothetical prote
28	6	13.3	107	2 F64488	hypothetical prote
29	6	13.3	108	2 C82864	plasmid maintenanc

RESULT 1

JH0557

exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi

N;Alternate names: neuraminidase

C;Species: Trypanosoma cruzi

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000

C;Accession: JH0557

R;Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.

J. Exp. Med. 174, 179-191, 1991

A;Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu

A;Reference number: JH0557; MUID:91277609; PMID:1711561

A;Accession: JH0557

A;Molecule type: DNA

A;Residues: 1-1162 <PER>

A;Cross-references: GB:M61732; NID:gl62302; PID:gl62303

A;Note: The authors translated the codon TCT for residue 45 as Cys

C;Comment: This protein plays a role in parasite-host cell interaction.

C;Superfamily: trypanastigote-specific surface antigen

C;Keywords: glycoprotein; glycosidase; hydrolase

F;394/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 62.2%; Score 28; DB 2; Length 1162;

Best Local Similarity 100.0%; Pred. No. 5.7e-20;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLRRQVVVVPLSPRLVLLAFRCRLP 28

Db 356 QPLRRQVVVVPLSPRLVLLAFRCRLP 383

RESULT 2

S09541

gene 1.6 protein - phase T3

C;Species: phase T3

C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 20-Sep-1999

C;Accession: S09541

R;Schmitt, M.P.; Beck, P.J.; Kearney, C.A.; Spence, J.L.; DiGiovanni, D.; Condreay, J.P.;

J. Mol. Biol. 193, 479-495, 1987

A;Title: Sequence of a conditionally essential region of bacteriophage T3, including the

A;Reference number: S07281; MUID:87226207; PMID:3586029

A;Accession: S09541

A;Molecule type: DNA

A;Residues: 1-85 <SCH>

A;Cross-references: EMBL:X05031; NID:gl5719; PIDN:CAA28702.1; PID:gl5726

C;Genetics:

A;Gene: 1.6

C;Superfamily: phase T7 gene 1.6 protein

Query Match 15.6%; Score 7; DB 2; Length 85;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 10 VVPLSPR 16
    |||||
Db 37 VVPLSPR 43

RESULT 3
F83473
hypothetical protein PA1378 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83473
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83473
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-120 <STO>
A;Cross-references: GB:AE004567; GB:AE004091; NID:9947316; PIDN:AG04767.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA1378

Query Match 15.6%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LSPRLVL 19
    |||||
Db 35 LSPRLVL 41

RESULT 4
B70964
hypothetical protein Rv2638 - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70964
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70964
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-148 <COL>
A;Cross-references: GB:Z80225; GB:AL123456; NID:g3242265; PIDN:CAB02344.1; PID:e266393;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2638

Query Match 15.6%; Score 7; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PLSPLRV 18
    |||||
Db 14 PLSPLRV 20

RESULT 5
AE2213
two-component response regulator alr3260 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AE2213

```

```

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:2195285; PMID:11759840
A;Accession: AE2213
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-238 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAE74959.1; PID:gl7132355; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3260
C;Superfamily: ompR protein; response regulator homology

Query Match 15.6%; Score 7; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRRQRVV 9
    |||||
Db 118 LRRQRVV 124

RESULT 6
AC3297
ABC transporter ATP-binding protein / ABC transporter permease protein BMEI0361 [imported]
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AC3297
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, N
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagijs, S.; O'Callaghan, D.; Letessc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AC3297
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51542.1; PID:gl7982260; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0361
A;Map position: 1
C;Superfamily: Bacillus subtilis conserved hypothetical protein yv6M

Query Match 15.6%; Score 7; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 VVVVPLS 14
    |||||
Db 71 VVVVPLS 77

RESULT 7
T30630
hypothetical protein 28L - Molluscum contagiosum virus 1
N;Alternate names: MC028L
C;Species: Molluscum contagiosum virus 1
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-May-2000
C;Accession: T30630
R;Senkevich, T.G.; Bugert, J.J.; Sisler, J.R.; Koonin, E.V.; Darai, G.; Moss, B.
Science 273, 813-816, 1996
A;Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host res
A;Reference number: Z20876; MUID:96325459; PMID:8670425
A;Accession: T30630
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-268 <SEN>
A;Cross-references: EMBL:U60315; PIDN:AAC55156.1
C;Genetics:
A;Note: MC028L

```

Query Match 15.6%; Score 7; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPLRROR 7
 Db 12 QPLRROR 18

RESULT 8
 AC2336
 potassium-dependent ATPase chain D' [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AC2336
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AC2336
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-377 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA075941.1; PID:gl7133377; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4242

Query Match 15.6%; Score 7; DB 2; Length 377;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VNASTAN 45
 Db 231 VNASTAN 237

RESULT 9
 A31589
 carboxypeptidase C (EC 3.4.16.5) precursor - human
 N:Alternate names: lysosomal protective protein
 C:Species: Homo sapiens (man)
 C:Date: 11-May-1989 #sequence_revision 11-May-1989 #text_change 05-May-2000
 C:Accession: A31589
 R:Galjart, N.J.; Gillenans, N.; Harris, A.; van der Horst, G.T.J.; Verheijen, F.W.; Galj
 Cell 54, 755-764, 1988
 A:Title: Expression of cDNA encoding the human "protective protein" associated with lysos
 A:Reference number: A31589; MUID:88311078; PMID:3136930
 A:Accession: A31589
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-480 <GAL>
 A:Cross-references: GB:M22960; GB:J03159; GB:M18453; PIDN:gl90282; PIDN:AAA36476.1; PID:9
 C:Superfamily: serine carboxypeptidase
 C:Keywords: hydrolase; serine carboxypeptidase
 F:178,400,457/Active site: Ser, Asp, His #status predicted

Query Match 15.6%; Score 7; DB 2; Length 480;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RLPLKRM 32
 Db 307 RLPLKRM 313

RESULT 10
 T44481
 aerobactin biosynthesis protein iucC [similarity] - Shigella flexneri
 C:Species: Shigella flexneri

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 16-Feb-2001
 C:Accession: T44481
 R:Moss, J.E.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
 Mol. Microbiol. 33, 74-83, 1999
 A:Title: The selC-associated SHI-2 pathogenicity island of Shigella flexneri.
 A:Reference number: 222779; MUID:99340340; PMID:10411725
 A:Accession: T44481
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-580 <MOS>
 A:Cross-references: EMBL:AF141323; NID:95532445; PIDN:AAD44748.1; PID:95532464
 A:Experimental source: strain M90T; serotype 5a
 C:Genetics:
 A:Gene: iucC
 C:Superfamily: Escherichia coli plasmid ColV-K30 aerobactin biosynthesis protein iucC

Query Match 15.6%; Score 7; DB 2; Length 580;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RVVVVPL 13
 Db 402 RVVVVPL 408

RESULT 11
 S50883
 aerobactin biosynthesis protein iucC [validated] - Escherichia coli plasmid ColV-K30
 C:Species: Escherichia coli
 C:Date: 07-May-1995 #sequence_revision 03-Oct-1995 #text_change 16-Feb-2001
 C:Accession: S50883; S44020
 R:Martinez, J.L.
 submitted to the EMBL Data Library, November 1993
 A:Reference number: S50883
 A:Accession: S50883
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-580 <MAR>
 A:Cross-references: EMBL:X76100; NID:9474189; PIDN:CAAS3709.1; PID:g474192
 R:Martinez, J.L.; Herrero, M.; de Lorenzo, V.
 J. Mol. Biol. 238, 288-293, 1994
 A:Title: The organization of intercastronic regions of the aerobactin operon of pColV-K30
 A:Reference number: S44019; MUID:94210503; PMID:8003107
 A:Accession: S44020
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-545 <MAR2>
 A:Cross-references: EMBL:X76100
 C:Genetics:
 A:Gene: iucC
 A:Genome: plasmid ColV-K30
 C:Function:
 A:Description: catalyzes the second step in aerobactin synthetase reaction [validated, M
 A:Pathway: aerobactin biosynthesis
 C:Superfamily: Escherichia coli plasmid ColV-K30 aerobactin biosynthesis protein iucC

Query Match 15.6%; Score 7; DB 2; Length 580;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RVVVVPL 13
 Db 402 RVVVVPL 408

RESULT 12
 E75407
 isoleucyl-tRNA synthetase - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75407
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75407
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1078 <WHI>
A;Cross-references: GB:AE001980; GB:AE000513; MID:g6459086; PIDN:AAF10907.1; PID:g645908
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRI335
A;Map position: 1
C;Superfamily: isoleucine-tRNA ligase

Query Match 15.6%; Score 7; DB 2; Length 1078;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 VPLSPR 16
| | | | |
Db 183 VPLSPR 189

RESULT 13
E39690
neural cell adhesion molecule, cardiac splice form -,+,+ - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 16-Jul-1999
C;Accession: E39690
R;Reyes, A.A.; Small, S.J.; Akeson, R.
Mol. Cell. Biol. 11, 1654-1661, 1991
A;Title: At least 27 alternatively spliced forms of the neural cell adhesion molecule mRNA
A;Reference number: A39690; MUID:91141516; PMID:1996115
A;Accession: E39690
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-24 <REY>
A;Cross-references: GB:M63970
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: cardiac muscle; cell adhesion; heart

Query Match 13.3%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VPLSPR 16
| | | | |
Db 13 VPLSPR 18

RESULT 14
G39690
neural cell adhesion molecule, cardiac splice form +,+,+ - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 16-Jul-1999
C;Accession: G39690
R;Reyes, A.A.; Small, S.J.; Akeson, R.
Mol. Cell. Biol. 11, 1654-1661, 1991
A;Title: At least 27 alternatively spliced forms of the neural cell adhesion molecule mRNA
A;Reference number: A39690; MUID:91141516; PMID:1996115
A;Accession: G39690
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-29 <REY>
A;Cross-references: GB:M63970
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: cardiac muscle; cell adhesion; heart

Query Match 13.3%; Score 6; DB 2; Length 29;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VPLSPR 16
| | | | |
Db 18 VPLSPR 23

RESULT 15
H39690
neural cell adhesion molecule, cardiac splice form -,+,+ - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 16-Jul-1999
C;Accession: H39690
R;Reyes, A.A.; Small, S.J.; Akeson, R.
Mol. Cell. Biol. 11, 1654-1661, 1991
A;Title: At least 27 alternatively spliced forms of the neural cell adhesion molecule mRNA
A;Reference number: A39690; MUID:91141516; PMID:1996115
A;Accession: H39690
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-38 <REY>
A;Cross-references: GB:M63970
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: cardiac muscle; cell adhesion; heart

Query Match 13.3%; Score 6; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VPLSPR 16
| | | | |
Db 13 VPLSPR 18

Search completed: March 4, 2003, 17:02:31
Job time : 20.8305 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:56:44 ; Search time 13.7288 Seconds
(without alignments)
135.950 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 45

Sequence: 1 QPLRRQVVVPLSPRLVLL.....RLPLKRMGGSYRCVNSTAN 45

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	62.2	1162	1 TCNA_TRYCR	P23253 trypanosoma
2	7	15.6	85	1 Y16_EPT3	P07718 bacterioph
3	7	15.6	480	1 PRTP_HUMAN	P10619 homo sapien
4	7	15.6	580	1 IUCC_ECOLI	Q47318 escherichia
5	6	13.3	69	1 V57_EPT3	P20320 bacterioph
6	6	13.3	85	1 DEFB_MOUSE	P50709 mus musculu
7	6	13.3	85	1 DEFB_MOUSE	P50712 mus musculu
8	6	13.3	88	1 Y4EA_RHISN	P55424 rhizobium s
9	6	13.3	91	1 DEFW_MOUSE	P50715 mus musculu
10	6	13.3	91	1 DEFW_MOUSE	P17534 mus musculu
11	6	13.3	92	1 DEFA_MOUSE	P28311 mus musculu
12	6	13.3	92	1 DEFA_MOUSE	P50708 mus musculu
13	6	13.3	93	1 DEFB_MOUSE	P28310 mus musculu
14	6	13.3	93	1 DEFB_MOUSE	P28312 mus musculu
15	6	13.3	93	1 DEFB_MOUSE	P50704 mus musculu
16	6	13.3	93	1 DEFB_MOUSE	P50705 mus musculu
17	6	13.3	93	1 DEFB_MOUSE	P50707 mus musculu
18	6	13.3	93	1 DEFB_MOUSE	P50711 mus musculu
19	6	13.3	93	1 DEFB_MOUSE	P50713 mus musculu
20	6	13.3	93	1 DEFB_MOUSE	P50714 mus musculu
21	6	13.3	94	1 Y054_ARCFU	O30182 archaeoglob
22	6	13.3	107	1 YF11_METJA	O58906 methanococ
23	6	13.3	109	1 TRP5_BOVIN	O9myv9 bos taurus
24	6	13.3	116	1 VMEM_WCMVM	P09500 white clove
25	6	13.3	129	1 PMVK_PIG	Q29081 sus scrofa
26	6	13.3	135	1 YEGB_RAT	O35485 rattus norv
27	6	13.3	136	1 TVC_HUMAN	P03979 homo sapien
28	6	13.3	147	1 IAA_HORVU	P16969 hordeum vul
29	6	13.3	167	1 EMP2_HUMAN	P54851 homo sapien
30	6	13.3	169	1 V55_EPT7	P03787 bacterioph
31	6	13.3	169	1 YBGA_ECOLI	P24252 escherichia
32	6	13.3	179	1 APT_HUMAN	P07741 homo sapien
33	6	13.3	191	1 PMVK_HUMAN	Q15126 homo sapien

34	6	13.3	206	1 R13A_PICMA	O65055 picea maria
35	6	13.3	207	1 YEGB_MOUSE	P49766 mus musculu
36	6	13.3	216	1 YGWB_YEAST	P53084 saccharomyc
37	6	13.3	223	1 COAT_CTV36	Q00686 citrus tris
38	6	13.3	228	1 ATPB_DROVI	Q24751 drosophila
39	6	13.3	234	1 HEN3_CHLPN	O925c9 chlamydia p
40	6	13.3	263	1 MTX2_HUMAN	O75431 homo sapien
41	6	13.3	263	1 MTX2_MOUSE	O88441 mus musculu
42	6	13.3	273	1 DAPB_ECO57	P58209 escherichia
43	6	13.3	273	1 DAPB_ECOLI	P04036 escherichia
44	6	13.3	273	1 DAPB_SALTI	Q82919 salmonella
45	6	13.3	273	1 DAPB_SALTY	Q829x8 salmonella

ALIGNMENTS

RESULT 1
TCNA_TRYCR STANDARD; PRT; 1162 AA.
AC P23253;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen).
GN TCNA.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Silvio X-10/4;
RX MEDLINE=91277609; PubMed=1711561;
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
RA Prioli R.P.;
RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
bacterial neuraminidases, YWTD repeats of the low density lipoprotein
receptor, and type III modules of fibronectin."
RL J. Exp. Med. 174:179-191(1991).
RN [2]
SUBCELLULAR LOCATION
MEDLINE=91376547; PubMed=1896773;
RA Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;
RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
trypanastigotes."
RL Trop. Med. Parasitol. 42:146-150(1991).
CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
PARASITE INVASION OF CELLS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
alpha-(2->8)-glycosidic linkages of terminal sialic residues in
oligosaccharides, glycoproteins, glycolipids, colominic acid and
synthetic substrates.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POSSIBLE).
CC -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
CC -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA
PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC -!- SIMILARITY: CONTAINS 3 BNR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M61732; AAA30255.1; -.
DR PIR; JH0557; JH0557.
DR InterPro; IPR002860; GH_BNR.

RX MEDLINE=96164441; PubMed=8591035;
RA Rudenko G., Bonten E., D'Azzo A., Hol W.G.J.;
RT "Three-dimensional structure of the human 'protective protein':
RT structure of the precursor form suggests a complex activation
RT mechanism.";
RL Structure 3:1249-1259 (1995).
RN [7]
RP VARIANT GALACTOSIALIDOSIS VAL-440.
RX MEDLINE=92097522; PubMed=1756715;
RA Zhou X.Y., Galjart N.J., Willemsen R., Gillemans N., Galjaard H.,
RA D'Azzo A.;
RT "A mutation in a mild form of galactosialidosis impairs dimerization
RT of the protective protein and renders it unstable.";
RL EMBO J. 10:4041-4048 (1991).
RN [8]
RP VARIANTS GALACTOSIALIDOSIS ARG-49; ARG-65; LEU-90; ASN-249 AND
RP CYS-395.
RX MEDLINE=93293970; PubMed=8514852;
RA Shimoto M., Fukuhara Y., Itoh K., Oshima A., Sakuraba H., Suzuki Y.;
RT "Protective protein gene mutations in galactosialidosis.";
RL J. Clin. Invest. 91:2393-2398 (1993).
CC -!- FUNCTION: PROTECTIVE PROTEIN APPEARS TO BE ESSENTIAL FOR BOTH THE
CC ACTIVITY OF BETA-GALACTOSIDASE AND NEURAMINIDASE, IT ASSOCIATES
CC WITH THESE ENZYMES AND EXERTS A PROTECTIVE FUNCTION NECESSARY FOR
CC THEIR STABILITY AND ACTIVITY. THIS PROTEIN IS ALSO A
CC CARBOXYPEPTIDASE AND CAN DEAMIDATE TACHYKININS.
CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -!- SUBUNIT: HETERODIMER OF A 32 kDa AND 20 kDa CHAINS HELD TOGETHER
CC BY DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- DISEASE: DEFECTS IN PPGB ARE THE CAUSE OF GALACTOSIALIDOSIS; AN
CC AUTOSOMAL RECESSIVE DISEASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22960; AAA36476.1; --
DR EMBL; AL008726; CAA15501.1; --
DR EMBL; BC000597; AAH00597.1; --
DR PIR; A31589; A31589.
DR PDB; 1IVY; 21-APR-97.
DR MEROPS; S10.002; --
DR Genew; HGNC:9251; PPGB.
DR MIM; 256540; --
DR InterPro; IPR000379; Ser esters site.
DR InterPro; IPR001563; Serine_carbpept.
DR Pfam; PF00450; serine_carbpept; 1.
DR PRINTS; PR00724; CRBOXYPTASEC.
DR ProDom; PD001189; Serine carbpept; 2.
DR PROSITE; PS00113; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal; Lysosome;
KW 3D-structure; Disease mutation.
FT SIGNAL 1 28
FT CHAIN 29 480 LYSOSOMAL PROTECTIVE PROTEIN.
FT CHAIN 29 326
FT CHAIN 32 326 KDA CHAIN.
FT CHAIN 327 480
FT ACT_SITE 178 178
FT ACT_SITE 400 400
FT ACT_SITE 457 457
FT DISULFID 88 362
FT DISULFID 240 256
FT DISULFID 241 246
FT DISULFID 281 331
FT DISULFID 145 145
FT CARBOHYD 333 333
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARIANT 49 49 Q -> R (IN GALACTOSIALIDOSIS).
FT /FTID=VAR 001385.
FT VARIANT 65 65 W -> R (IN GALACTOSIALIDOSIS).
FT /FTID=VAR 001386.
FT VARIANT 90 90 S -> L (IN GALACTOSIALIDOSIS).
FT /FTID=VAR 001387.
FT VARIANT 249 249 Y -> N (IN GALACTOSIALIDOSIS; SMALL
FT AMOUNT OF ACTIVITY).
FT /FTID=VAR 001388.
FT VARIANT 395 395 Y -> C (IN GALACTOSIALIDOSIS; LOSS
FT OF ACTIVITY).
FT /FTID=VAR 001389.
FT VARIANT 440 440 F -> V (IN GALACTOSIALIDOSIS).
FT /FTID=VAR 001390.
FT MUTAGEN 178 178 S->A: INACTIVATES ENZYME.
FT MUTAGEN 457 457 H->Q: INACTIVATES ENZYME.
FT CONFLICT 19 19 MISSING (IN REF. 3).
FT CONFLICT 56 56 G -> S (IN REF. 1).
SQ SEQUENCE 480 AA; 54466 MW; 46B737DEE775C508 CRC64;
Query Match 15.6%; Score 7; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RLPLKRM 32
Db 307 RLPLKRM 313
RESULT 4
ID IUCC_ECOLI STANDARD; PRT; 580 AA.
AC Q47318;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Aerobactin siderophore biosynthesis protein iucc.
GN IUCC.
OS Escherichia coli.
OG Plasmid IncPI ColV3-K30.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94210503; PubMed=8003107;
RA Martinez J.L., Herrero M., de Lorenzo V.;
RT "The organization of intergenic regions of the aerobactin operon
RT of pColV-K30 may account for the differential expression of the
RT iucABCD iutA genes.";
RL J. Mol. Biol. 238:288-293 (1994).
CC -!- PATHWAY: Aerobactin biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE IUCC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X76100; CAA53709.1; --
DR Iron transport; Plasmid.
SQ SEQUENCE 580 AA; 66845 MW; 406FB857BB367F65 CRC64;
Query Match 15.6%; Score 7; DB 1; Length 580;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 RVVVVPL 13
Db 7 RVVVVPL 13


```
CC -!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL BOWEL.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U03067; AAA03714.1; -.
DR MGD; MGI:99587; Defcrl4
DR InterPro; IPR002366; Defensin_alpha.
DR InterPro; IPR001271; Defensin_mammal.
DR Pfam; PF00323; defensins; 1.
DR Pfam; PF00879; Defensin_propep; 1.
DR SMART; SM00048; DEFSN; 1.
DR PROSITE; PS00269; DEFENSIN; 1.
KW Defensin; Antibiotic; Signal.
FT NON_TER 1 1
FT SIGNAL <1 11 POTENTIAL.
FT PROPEP 12 50 BY SIMILARITY.
FT CHAIN 51 85 CRYPTDIN-14.
FT DISULFID 56 84 BY SIMILARITY.
FT DISULFID 58 73 BY SIMILARITY.
FT DISULFID 63 83 BY SIMILARITY.
SQ SEQUENCE 85 AA; 9589 MW; 78F4A9011AF7D13A CRC64;

Query Match 13.3%; Score 6; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
DB 2 LVLLAF 7

RESULT 8
ID Y4EA RHISN STANDARD; PRT; 88 AA.
AC P55424;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 9.9 kDa protein Y4EA.
GN Y4EA
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A5000070; AAB92445.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 88 AA; 9896 MW; 704EFP28C51B5F3C CRC64;
```

```
Query Match 13.3%; Score 6; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 PRVLVL 20
DB 37 PRVLVL 42

RESULT 9
ID DEFW MOUSE STANDARD; PRT; 91 AA.
AC P50715; Q64109;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cryptdin-related protein 4C-2 precursor (CRS4C).
GN DEFCR-RS7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ; TISSUE=Small intestine;
RX MEDLINE=95203896; PubMed=7896294;
RA Huttner K.M., Ouellette A.J.;
RT "A family of defensin-like genes codes for diverse cysteine-rich
RT peptides in mouse Paneth cells.";
RL Genomics 24:99-109(1994).
CC -!- FUNCTION: CRS4C IS AN APPARENT PRECURSOR OF A SECRETED, CATIONIC,
CC PROLINE- AND CYSTEINE-RICH PEPTIDE THAT CONTAINS CYS-PRO-XAA
CC REPEATS. UNLIKE CRYPTDIN, THE PROPOSED CRS4C MATURE PEPTIDE REGION
CC LACKS THE STRUCTURAL MOTIF CHARACTERISTIC OF DEFENSINS. IT MAY
CC HAVE MICROBICIDAL ACTIVITIES.
CC -!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL BOWEL.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S77610; AAB33826.2; -.
DR EMBL: S77606; AAB33826.2; JOINED.
DR EMBL: U12564; AAA20977.1; ALT_SEQ.
DR MGD; MGI:102509; Defcr-rs7.
DR InterPro; IPR002366; Defensin_alpha.
DR Pfam; PF00879; Defensin_propep; 1.
KW Defensin; Repeat; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 65 POTENTIAL.
FT CHAIN 66 91 CRYPTDIN-RELATED PROTEIN 4C-2.
FT DOMAIN 65 85 6 X 3 AA TANDEM REPEATS OF C-P-X.
FT REPEAT 65 67 1.
FT REPEAT 68 70 2.
FT REPEAT 71 73 3.
FT REPEAT 77 79 4.
FT REPEAT 80 82 5.
FT REPEAT 83 85 6.
SQ SEQUENCE 91 AA; 10055 MW; B285679960C376D4 CRC64;

Query Match 13.3%; Score 6; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
DB 10 LVLLAF 15
```



```

DR SMART; SM00048; DEFSN; 1.
DR PROSITE; PS00269; DEFENSIN; FALSE_NEG.
KW Defensin; Antibiotic; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 58
FT CHAIN 59 92 CRYPTDIN-4.
FT DISULFID 64 89 BY SIMILARITY.
FT DISULFID 66 81 BY SIMILARITY.
FT DISULFID 71 88 BY SIMILARITY.
SQ SEQUENCE 92 AA; 10272 MW; BF632838D52AFCE2 CRC64;

Query Match 13.3%; Score 6; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
DB 10 LVLLAF 15

RESULT 12
DEFA_MOUSE STANDARD; PRT; 92 AA.
AC P50708;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cryptdin-10 precursor (Fragment).
GN DEFCR10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Intestinal crypts;
RX MEDLINE=95012724; PubMed=7927786;
RA Ouellette A.J., Heieh M.M., Nosek M.T., Cano-Gauci D.F.,
RA Huttner K.M., Buick R.N., Selsted M.E.;
RT "Mouse Paneth cell defensins; primary structures and antibacterial
RT activities of numerous cryptdin isoforms.";
RL Infect. Immun. 62:5040-5047(1994).
RN [2]
RP SEQUENCE OF 58-92 FROM N.A.
RC STRAIN=129/SVJ, and C3H/HeJ; TISSUE=Small intestine;
RX MEDLINE=94245232; PubMed=8188287;
RA Huttner K.M., Selsted M.E., Ouellette A.J.;
RT "Structure and diversity of the murine cryptdin gene family.";
RL Genomics 19:448-453(1994).
CC -!- FUNCTION: PROBABLY CONTRIBUTES TO THE ANTIMICROBIAL BARRIER
CC -!- FUNCTION OF THE SMALL BOWEL MUCOSA.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03061; AAA57178.1; -
DR MGD; MGI:99591; Defcrl0.
DR InterPro; IPR002366; Defensin_alpha.
DR InterPro; IPR001271; Defensin_mammal.
DR Pfam; PF00323; defensins; 1.
DR Pfam; PF00879; Defensin_propep; 1.
DR SMART; SM00048; DEFSN; 1.
DR PROSITE; PS00269; DEFENSIN; 1.
KW Defensin; Antibiotic; Signal.
FT NON_TER 1 1
FT SIGNAL <1 18 POTENTIAL.

FT SMART; SM00048; DEFSN; 1.
FT PROSITE; PS00269; DEFENSIN; FALSE_NEG.
KW Defensin; Antibiotic; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 58
FT CHAIN 59 92 CRYPTDIN-4.
FT DISULFID 64 89 BY SIMILARITY.
FT DISULFID 66 81 BY SIMILARITY.
FT DISULFID 71 88 BY SIMILARITY.
SQ SEQUENCE 92 AA; 10298 MW; E2BD983E2DDC1E50 CRC64;

Query Match 13.3%; Score 6; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
DB 9 LVLLAF 14

RESULT 13
DEF3_MOUSE STANDARD; PRT; 93 AA.
AC P28310;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cryptdin-3 precursor.
GN DEFCR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Small intestine;
RX MEDLINE=94245232; PubMed=8188287;
RA Huttner K.M., Selsted M.E., Ouellette A.J.;
RT "Structure and diversity of the murine cryptdin gene family.";
RL Genomics 19:448-453(1994).
RN [2]
RP SEQUENCE OF 59-93.
RC STRAIN=Swiss; TISSUE=Small intestine;
RX MEDLINE=92363933; PubMed=1500431;
RA Selsted M.E., Miller S.I., Henschen A.H., Ouellette A.J.;
RT "Enteric defensins: antibiotic peptide components of intestinal host
RT defense.";
RL J. Cell Biol. 118:929-936(1992).
CC -!- FUNCTION: PROBABLY CONTRIBUTES TO THE ANTIMICROBIAL BARRIER
CC -!- FUNCTION OF THE SMALL BOWEL MUCOSA.
CC -!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL BOWEL.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12559; AAA20972.1; -
DR EMBL; U02999; AAB60677.1; -
DR EMBL; U02998; AAB60677.1; JOINED.
DR EMBL; U03030; AAA57171.1; -
DR PIR; C43279; C43279.
DR MGD; MGI:94883; Defcrl3.
DR InterPro; IPR002366; Defensin_alpha.
DR InterPro; IPR001271; Defensin_mammal.
DR Pfam; PF00323; defensins; 1.
DR Pfam; PF00879; Defensin_propep; 1.
DR SMART; SM00048; DEFSN; 1.
DR PROSITE; PS00269; DEFENSIN; 1.
KW Defensin; Antibiotic; Signal.
FT SIGNAL 1 16
FT PROPEP 17 58
FT CHAIN 59 93 CRYPTDIN-3.

```

```
FT DISULFID 64 92 BY SIMILARITY.
FT DISULFID 66 81 BY SIMILARITY.
FT DISULFID 71 91 BY SIMILARITY.
SQ SEQUENCE 93 AA; 10528 MW; 2B2D61CFCF40E634 CRC64;
Query Match 13.3%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LVLLAF 22
Db 10 LVLLAF 15

RESULT 14
DEF5_MOUSE
ID DEF5_MOUSE STANDARD; PRT; 93 AA.
AC P28312; Q64382;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cryptidin-5 precursor.
GN DEFCR5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (DEF5_MOUSE), AND PARTIAL SEQUENCE FROM N.A. (DEF5_MOUSE).
RC STRAIN=129, and C3H/HeJ; TISSUE=Intestinal epithelium;
RX MEDLINE=94245232; PubMed=8188287;
RA Huttner K.M., Selsted M.E., Ouellette A.J.;
RT "Structure and diversity of the murine cryptidin gene family.";
RL Genomics 19:448-453(1994).
RN [2]
RP SEQUENCE OF 59-93.
RC STRAIN=Swiss; TISSUE=Small intestine;
RX MEDLINE=92363933; PubMed=1500431;
RA Selsted M.E., Miller S.I., Henschen A.H., Ouellette A.J.;
RT "Enteric defensins: antibiotic peptide components of intestinal host defense.";
RL J. Cell Biol. 118:929-936(1992).
CC -1- FUNCTION: PROBABLY CONTRIBUTES TO THE ANTIMICROBIAL BARRIER
CC -1- FUNCTION OF THE SMALL BOWEL MUCOSA.
CC -1- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U12560; AAA20973.1; --
CC ENBL; U03001; AAB60678.1; --
CC ENBL; U03000; AAB60678.1; JOINED.
CC ENBL; U03033; AAA57173.1; --
CC MGD; MGI:99583; Defcr5.
CC InterPro; IPR002366; Defensin_alpha.
CC InterPro; IPR001271; Defensin_mammal.
CC Pfam; PF00323; defensins; 1.
CC Pfam; PF00879; Defensin_propep; 1.
CC SMART; SM00048; DEFSN; 1.
CC PROSITE; PS00269; DEFENSIN; 1.
KW Defensin; Antibiotic; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 58
FT CHAIN 59 93 CRYPTIDIN-5.
FT BY SIMILARITY.
FT DISULFID 64 92 BY SIMILARITY.
FT DISULFID 66 81 BY SIMILARITY.
FT DISULFID 71 91 BY SIMILARITY.
FT CONFLICT 59 59 S -> SA (IN REF. 2).
```

```
SQ SEQUENCE 93 AA; 10519 MW; 7FC27CBBFB6259E CRC64;
Query Match 13.3%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 LVLLAF 22
Db 10 LVLLAF 15

RESULT 15
DEF6_MOUSE
ID DEF6_MOUSE STANDARD; PRT; 93 AA.
AC P50704; P50710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cryptidin-6/12 precursor.
GN DEFCR6 AND DEFCR12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (DEF6_MOUSE), AND PARTIAL SEQUENCE FROM N.A. (DEF6_MOUSE).
RC STRAIN=CD-1; TISSUE=Intestinal crypts;
RX MEDLINE=95012724; PubMed=7927786;
RA Ouellette A.J., Heieh M.M., Nosek M.T., Cano-Gauci D.F.,
RA Huttner K.M., Buick R.N., Selsted M.E.;
RT "Mouse Paneth cell defensins: primary structures and antibacterial activities of numerous cryptidin isoforms.";
RL Infect. Immun. 62:5040-5047(1994).
CC -1- FUNCTION: PROBABLY CONTRIBUTES TO THE ANTIMICROBIAL BARRIER
CC -1- FUNCTION OF THE SMALL BOWEL MUCOSA.
CC -1- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL BOWEL.
CC -1- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; U12561; AAA20974.1; --
CC ENBL; U03003; AAB60679.1; --
CC ENBL; U03002; AAB60679.1; JOINED.
CC ENBL; U03034; AAA57174.1; --
CC ENBL; U03063; AAA57180.1; --
CC MGD; MGI:99582; Defcr6.
CC MGD; MGI:99589; Defcr12.
CC InterPro; IPR002366; Defensin_alpha.
CC InterPro; IPR001271; Defensin_mammal.
CC Pfam; PF00323; defensins; 1.
CC Pfam; PF00879; Defensin_propep; 1.
CC SMART; SM00048; DEFSN; 1.
CC PROSITE; PS00269; DEFENSIN; 1.
KW Defensin; Antibiotic; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 60 BY SIMILARITY.
FT CHAIN 61 93 CRYPTIDIN-6/12.
FT BY SIMILARITY.
FT DISULFID 64 92 BY SIMILARITY.
FT DISULFID 66 81 BY SIMILARITY.
FT DISULFID 71 91 BY SIMILARITY.
```

FT VARIANT 61 61 D -> Y (IN STRAIN C3H/HEJ) .
FT VARIANT 87 87 L -> M (IN STRAIN C3H/HEJ) .
FT VARIANT 89 89 M -> T (IN STRAIN C3H/HEJ) .
SQ SEQUENCE 93 AA; 10426 MW; 84CE684E476FB885 CRC64;

Query Match 13.3%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
| | | | |
Db 10 LVLLAF 15

Search completed: March 4, 2003, 17:00:31
Job time : 14.7288 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:57:01 ; Search time 37.3729 Seconds
(without alignments)
248.098 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 45
Sequence: 1 QPLRRQVVVPLSPRLVLL.....RLPLKRMGSGYRCVNSTAN 45

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: SPTREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phase.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 18: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	10	22.2	642 5	Q26966 trypanosoma
2	10	22.2	642 5	Q26967 trypanosoma
3	10	22.2	642 5	Q26968 trypanosoma
4	10	22.2	643 5	Q9BHJ5 trypanosoma
5	10	22.2	700 5	Q26965 trypanosoma
6	10	22.2	736 5	Q26850 trypanosoma
7	10	22.2	964 5	Q26963 trypanosoma
8	10	22.2	1060 5	Q26964 trypanosoma
9	7	15.6	85 9	Q9TI38 bacterioph
10	7	15.6	120 16	Q9I3W6 pseudomonas
11	7	15.6	143 5	Q9NG17 haemochus
12	7	15.6	148 16	P71937 mycobacteri
13	7	15.6	192 11	Q9DI92 mus musculu
14	7	15.6	210 16	Q92LB6 rhizobium m
15	7	15.6	231 5	Q9VXS3 drosophila
16	7	15.6	238 16	Q8YS31 anabaena sp

17	7	15.6	266	16	Q8YIT1	O8YIT1 brucella me
18	7	15.6	268	12	Q98196	Q98196 molluscum c
19	7	15.6	269	2	Q9WMM2	Q9WMM2 synechococc
20	7	15.6	281	17	Q979W3	Q979W3 thermoplasm
21	7	15.6	377	16	Q8YPF2	O8YPF2 anabaena sp
22	7	15.6	386	4	O75086	O75086 homo sapien
23	7	15.6	410	5	Q9VXS4	Q9VXS4 drosophila
24	7	15.6	443	2	Q9XST4	Q9XST4 streptomyce
25	7	15.6	518	8	Q9GHE3	Q9GHE3 drymophila
26	7	15.6	548	4	Q9BR08	Q9BR08 homo sapien
27	7	15.6	580	2	Q9XCH2	Q9XCH2 shigella fl
28	7	15.6	580	2	Q9JPE4	Q9JPE4 shigella bo
29	7	15.6	642	5	Q26969	Q26969 trypanosoma
30	7	15.6	1078	16	Q9RUP8	Q9RUP8 deinococcus
31	7	15.6	1092	16	Q98FB4	Q98FB4 rhizobium l
32	7	15.6	1322	5	Q9NJS5	Q9NJS5 anopheles g
33	7	15.6	1322	5	Q9NAT0	Q9NAT0 anopheles g
34	7	15.6	1396	3	Q8WZY9	Q8WZY9 neurospora
35	6	13.3	15	11	O88175	O88175 mus musculu
36	6	13.3	37	11	Q04017	Q04017 rattus norv
37	6	13.3	37	12	Q9J473	Q9J473 citrus tris
38	6	13.3	49	11	Q61947	Q61947 mus musculu
39	6	13.3	55	16	Q8YQ57	Q8YQ57 anabaena sp
40	6	13.3	69	9	Q9T118	Q9T118 bacterioph
41	6	13.3	69	13	Q9I214	Q9I214 oncorhynch
42	6	13.3	69	16	Q9PE34	Q9PE34 xyella fas
43	6	13.3	70	12	Q994E1	Q994E1 porcine ade
44	6	13.3	73	12	Q91M35	Q91M35 citrus tris
45	6	13.3	73	12	Q91M34	Q91M34 citrus tris

ALIGNMENTS

RESULT 1

Q26966 PRELIMINARY; PRT; 642 AA.
AC Q26966;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021510; PubMed=7935611;
RA Campetella O.E., Uttaro A.D., Parodi A.J., Frasch A.C.;
RT "A recombinant Trypanosoma cruzi trans-sialidase lacking the amino
acid repeats retains the enzymatic activity.";
RL Mol. Biochem. Parasitol. 64:337-340(1994).
DR EMBL; L26499; AAA66352.1; -;
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70593 MW; 65BD95DD7ADBC222 CRC64;

Query Match 22.2%; Score 10; DB 5; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNSTAN 45

Db 437 YRCVNSTAN 446

RESULT 2

Q26967 PRELIMINARY; PRT; 642 AA.
ID Q26967
AC Q26967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

```
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RA;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RL Gene 160:123-128(1995).";
DR EMBL; L38456; AAA99442.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70453 MW; BC263FAAAF61EAGE CRC64;

Query Match 22.2%; Score 10; DB 5; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 437 YRCVNASTAN 446
|||||

RESULT 3
Q26968 PRELIMINARY; PRT; 642 AA.
AC Q26968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TULAHUEN;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RL Gene 160:123-128(1995).";
DR EMBL; L38457; AAA99443.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 642 AA; 70459 MW; 09F047336492081B CRC64;

Query Match 22.2%; Score 10; DB 5; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 437 YRCVNASTAN 446
|||||

RESULT 4
Q9BHJ5 PRELIMINARY; PRT; 643 AA.
AC Q9BHJ5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=Y-STRAIN;
RA Laroy W., Contreras R.;
RT "Cloning of Trypanosoma cruzi trans-sialidase and expression in Pichia
RL pastoris.";
DR EMBL; AJ276679; CAC34453.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 643 AA; 70546 MW; E46D9A7CE9218ECB CRC64;

Query Match 22.2%; Score 10; DB 5; Length 643;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 438 YRCVNASTAN 447
|||||

RESULT 5
Q26965 PRELIMINARY; PRT; 700 AA.
AC Q26965;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase homolog (Fragment).
GN TCTS-193.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
RL from Trypanosoma rangeli.";
DR EMBL; D50686; BAA09335.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 2.
FT NON TER 700
SQ SEQUENCE 700 AA; 77107 MW; B5E6CB448261796B CRC64;

Query Match 22.2%; Score 10; DB 5; Length 700;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 461 YRCVNASTAN 470
|||||

RESULT 6
Q26850 PRELIMINARY; PRT; 736 AA.
AC Q26850;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase homologue.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=95364855; PubMed=7637718;
RA Briones M.R.S., Egima C.M., Schenkman S.;
RT "Trypanosoma cruzi trans-sialidase gene lacking C-terminal repeats and
RL expressed in epimastigote forms.";
DR EMBL; U00000; U00000.1; -.
DR Mol. Biochem. Parasitol. 70:9-17(1995).
```

```

[2]
RN SEQUENCE OF 124-324 FROM N.A.
RC STRAIN=Y;
RA MEDLINE=95395867; PubMed=7666441;
RA Briones M.R.S., Egima C.M., Eichinger D., Schenkman S.;
RT "Trans-sialidase genes expressed in mammalian forms of Trypanosoma
RT cruzi evolved from ancestor genes expressed in insect forms of the
RT parasite.";
RL J. Mol. Evol. 41:120-131 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Briones M.R.S.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; U01098; AAC98544.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 2.
SQ SEQUENCE 736 AA; 81997 MW; 75A8026CE8D7795A CRC64;

Query Match 22.2%; Score 10; DB 5; Length 736;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 461 YRCVNASTAN 470
|||||

RESULT 7
Q26963 PRELIMINARY; PRT; 964 AA.
AC Q26963;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trans-sialidase.
GN TCTS-121.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Uemura H.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 178-378 FROM N.A.
RC STRAIN=Y;
RX MEDLINE=93010978; PubMed=1396577;
RA Smith L.E., Uemura H., Nussenzweig V., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
RT from Trypanosoma rangeli.";
RL Mol. Biochem. Parasitol. 79:21-34 (1996).
DR EMBL; D50685; BAA09334.1; -.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 3.
SQ SEQUENCE 1060 AA; 110103 MW; EBD146AD6DD26288 CRC64;

Query Match 22.2%; Score 10; DB 5; Length 1060;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
Db 468 YRCVNASTAN 477
|||||

RESULT 9
Q9T138 PRELIMINARY; PRT; 85 AA.
AC Q9T138;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19; Last annotation update)
DE 1.6 protein.
GN 1.6.
OS Bacteriophage phiYeO3-12.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=110457;
RN [1]
RP SEQUENCE FROM N.A.
RA Pajunen M.I.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21125554; PubMed=11222590;
RA Pajunen M.I., Kiljunen S.J., Soederholm M.E.L., Skurnik M.;
RT "Complete genomic sequence of the lytic bacteriophage phiYeO3-12 of
```

```

RT Yersinia enterocolitica serotype O:3."
RL J. Bacteriol. 183:1928-1937(2001).
DR EMBL: AJ251805; CAB63599.1; -.
SQ SEQUENCE 85 AA; 9894 MW; 2FSD0D8628C150BE CRC64;

Query Match 15.6%; Score 7; DB 9; Length 85;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 VVPLSPR 16
Db 37 VVPLSPR 43
|||||

RESULT 10
Q913W6 PRELIMINARY; PRT; 120 AA.
AC Q913W6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein PA1378.
GN PA1378.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goulet S., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
RW EMBL: AE004567; AAG04767.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 13656 MW; 1EBD55456A4D6632 CRC64;

Query Match 15.6%; Score 7; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 LSPRLVL 19
Db 35 LSPRLVL 41
|||||

RESULT 11
Q9NG17 PRELIMINARY; PRT; 143 AA.
AC Q9NG17;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Galectin.
GN GAL16.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RA Schmitt-Wrede H.P., Blasum H., Saeger B., Harder A., Wunderlich F.;
RT "Proto-type galectin from intestinal nematode Haemonchus contortus."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ243873; CAB71314.1; -.
DR HSSP; P47929; 2GAL.

Query Match 15.6%; Score 7; DB 16; Length 148;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PLSPRLV 18
Db 14 PLSPRLV 20
|||||

DR InterPro; IPR001079; Galectin.
DR SMART; PF00337; Gal-bind lectin; 1.
DR SMART; SM00276; GLECT; 1.
SQ SEQUENCE 143 AA; 16294 MW; 42F6D39B2A1733F6 CRC64;

Query Match 15.6%; Score 7; DB 5; Length 143;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RVVVVPL 13
Db 33 RVVVVPL 39
|||||

RESULT 12
P71937 PRELIMINARY; PRT; 148 AA.
AC P71937;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SPOL1AA family protein.
GN RV2638 OR MTCY441.08 OR MT2716.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z80225; CAB02344.1; -.
DR TIGR; AE007103; AAK47030.1; -.
DR TIGR; MT2716; -.
DR TubercuList; RV2638; -.
DR InterPro; IPR003658; Antisig_antgst.
DR InterPro; IPR002645; STAS.
DR PFam; PF01740; STAS; 1.
DR TIGRFAMs; TIGR00377; ant ant sig; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 15419 MW; C279BC99E1ABACE7 CRC64;

Query Match 15.6%; Score 7; DB 16; Length 148;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PLSPRLV 18
Db 14 PLSPRLV 20
|||||

```

RESULT 13

Q9D1G2 ID Q9D1G2 PRELIMINARY; PRT; 192 AA.
 AC Q9D1G2;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE 1110011E12RIK protein.
 GN 1110011E12RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaldi P.,
 RA Nordone P., Ring B., Ringwald C., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.
 RA "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL EMBL; AK003607; BAB22886.1; --
 DR MGD; MGI:1915853; 1110011E12RIK.
 DR TIGRFAMs; TIGR01223; Pmev kin anim; 1.
 SQ SEQUENCE 192 AA; 21932 MW; F0715558C2273E33A CRC64;

Query Match 15.6%; Score 7; DB 11; Length 192;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 SPRLVLL 20
 Db 7 SPRLVLL 13

RESULT 14

Q92LB6 ID Q92LB6 PRELIMINARY; PRT; 210 AA.
 AC Q92LB6;
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
 DE Hypothetical protein R03157.
 GN R03157 OR SMC03773.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subphylum; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Guzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.

RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591793; CAC47736.1; --
 DR InterPro; IPR000182; GCNSacetyltransf.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 210 AA; 23545 MW; 1FF6480A8043FF9 CRC64;

Query Match 15.6%; Score 7; DB 16; Length 210;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 LSPRLVL 19
 Db 31 LSPRLVL 37

RESULT 15

Q9VXS3 ID Q9VXS3 PRELIMINARY; PRT; 231 AA.
 AC Q9VXS3;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last annotation update)
 DE CG12708 protein.
 GN CG12708.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.D.,
 RA Abell R.M., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
 RA Baller J.F., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beran B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."

```

RL Science 287:2185-2195 (2000).
DR EMBL; AE003499; AAF48485.1; -.
DR FlyBase; Fggn0030666; CG12708.
SQ SEQUENCE 231 AA; 26408 MW; 5A3D911767A9860E CRC64;

Query Match 15.6%; Score 7; DB 5; Length 231;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels

Qy 3 LRQRVV 9
      |||||
Db 122 LRQRVV 128

```

Search completed: March 4, 2003, 17:01:28
Job time : 39.3729 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:58:06 ; Search time 17.524 Seconds
(without alignments)
75.476 Million cell updates/sec

Title: US-09-745-008-12
Perfect score: 45
Sequence: 1 QPLRRQVVVPLSPRLVLL.....RLPLKRMGGSYRCVNSTAN 45

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pdp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	22.2	642	4	US-08-911-393-4
2	10	22.2	1060	4	US-08-911-393-2
3	7	15.6	42	2	US-08-766-858A-32
4	7	15.6	480	2	US-08-828-488-8
5	7	15.6	480	4	US-09-299-689A-8
6	6	13.3	21	3	US-08-746-111-21
7	6	13.3	85	1	US-08-486-013-33
8	6	13.3	85	1	US-08-486-013-36
9	6	13.3	85	2	US-08-482-279-33
10	6	13.3	85	2	US-08-482-279-36
11	6	13.3	85	2	US-08-342-268-33
12	6	13.3	85	2	US-08-342-268-36
13	6	13.3	85	3	US-09-015-968-33
14	6	13.3	85	3	US-09-015-968-36
15	6	13.3	85	4	US-09-397-386-33
16	6	13.3	85	4	US-09-397-386-36
17	6	13.3	92	1	US-08-486-013-26
18	6	13.3	92	1	US-08-486-013-32
19	6	13.3	92	2	US-08-482-279-26
20	6	13.3	92	2	US-08-482-279-32
21	6	13.3	92	2	US-08-342-268-26
22	6	13.3	92	2	US-08-342-268-32
23	6	13.3	92	3	US-09-015-968-26
24	6	13.3	92	3	US-09-015-968-32
25	6	13.3	92	4	US-09-397-386-26
26	6	13.3	92	4	US-09-397-386-32
27	6	13.3	92	4	US-09-257-179-42

28	6	13.3	93	1	US-08-486-013-23	Sequence 23, Appl
29	6	13.3	93	1	US-08-486-013-24	Sequence 24, Appl
30	6	13.3	93	1	US-08-486-013-25	Sequence 25, Appl
31	6	13.3	93	1	US-08-486-013-27	Sequence 27, Appl
32	6	13.3	93	1	US-08-486-013-28	Sequence 28, Appl
33	6	13.3	93	1	US-08-486-013-29	Sequence 29, Appl
34	6	13.3	93	1	US-08-486-013-30	Sequence 30, Appl
35	6	13.3	93	1	US-08-486-013-34	Sequence 34, Appl
36	6	13.3	93	1	US-08-486-013-35	Sequence 35, Appl
37	6	13.3	93	1	US-08-486-013-37	Sequence 37, Appl
38	6	13.3	93	1	US-08-486-013-38	Sequence 38, Appl
39	6	13.3	93	2	US-08-482-279-23	Sequence 23, Appl
40	6	13.3	93	2	US-08-482-279-24	Sequence 24, Appl
41	6	13.3	93	2	US-08-482-279-25	Sequence 25, Appl
42	6	13.3	93	2	US-08-482-279-27	Sequence 27, Appl
43	6	13.3	93	2	US-08-482-279-28	Sequence 28, Appl
44	6	13.3	93	2	US-08-482-279-29	Sequence 29, Appl
45	6	13.3	93	2	US-08-482-279-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-08-911-393-4
; Sequence 4, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: SIALYLOIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 642 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-911-393-4

Query Match 22.2%; Score 10; DB 4; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 YRCVNSTAN 45
| | | | | | | | | |
Db 437 YRCVNSTAN 446

RESULT 2

US-08-911-393-2
; Sequence 2, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; STALYLOLIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-911-393-2

Query Match 22.2%; Score 10; DB 4; Length 1060;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 YRCVNSTAN 45
| | | | | | | | | |
Db 468 YRCVNSTAN 477

RESULT 3

US-08-766-858A-32
; Sequence 32, Application US/08766858A
; Patent No. 5935782
; GENERAL INFORMATION:
; APPLICANT: Cianciotto, Nicholas P.
; APPLICANT: Hickey, Erin K.
; APPLICANT: O'Connell, William A.
; TITLE OF INVENTION: METHODS AND MATERIALS FOR DETECTING

; TITLE OF INVENTION: LEGIONELLA PNEUMOPHILA
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,858A
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,545
; FILING DATE: 13-FEB-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Crook, Wannell M.

; REGISTRATION NUMBER: 31,071

; REFERENCE/DOCKET NUMBER: 3501-4-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 32:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 42 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-766-858A-32

Query Match 15.6%; Score 7; DB 2; Length 42;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RVVVVPL 13
| | | | | | | |
Db 32 RVVVVPL 38

RESULT 4

US-08-828-488-8

; Sequence 8, Application US/08828488

; Patent No. 5925521

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; TITLE OF INVENTION: NOVEL HUMAN SERINE

; TITLE OF INVENTION: CARBOXYPEPTIDASE

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/828,488


```

; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 190283
; US-08-828-488-8

Query Match 15.6%; Score 7; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RLPLKRM 32
Db 307 RLPLKRM 313

RESULT 5
US-09-299-689A-8
; Sequence 8, Application US/09299689A
; Patent No. 6378913
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN SERINE
; TITLE OF INVENTION: CARBOXYPEPTIDASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,689A
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,488
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0241 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 190283
; US-09-299-689A-8

Query Match 15.6%; Score 7; DB 4; Length 480;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RLPLKRM 32
Db 307 RLPLKRM 313

RESULT 6
US-08-746-111-21
; Sequence 21, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-746-111-21

Query Match 13.3%; Score 6; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 VPLSPR 16
Db 6 VPLSPR 11

RESULT 7
US-08-486-013-33
; Sequence 33, Application US/08486013
; Patent No. 5731149
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.

```

APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/486,013
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-486-013-33

Query Match 13.3%; Score 6; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 8
US-08-486-013-36
Sequence 36, Application US/08486013
Patent No. 5731149
GENERAL INFORMATION:
APPLICANT: Seistred, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,013
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-486-013-36
Query Match 13.3%; Score 6; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 9
US-08-482-279-33
Sequence 33, Application US/08482279
Patent No. 5840498
GENERAL INFORMATION:
APPLICANT: Seistred, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,279
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-482-279-33

Query Match 13.3%; Score 6; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 10
US-08-482-279-36
Sequence 36, Application US/08482279
Patent No. 5840498
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,279
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/342,268
FILING DATE: 18-NOV-1994
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-482-279-36

Query Match 13.3%; Score 6; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
Db 2 LVLLAF 7

Db 2 LVLLAF 7

RESULT 11
US-08-342-268-33
Sequence 33, Application US/08342268
Patent No. 5844072
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,268
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,649
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/889,020
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UC 1206
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-342-268-33

Query Match 13.3%; Score 6; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 12
US-08-342-268-36
Sequence 36, Application US/08342268
Patent No. 5844072
GENERAL INFORMATION:
APPLICANT: Selsted, Michael E.
APPLICANT: Ouellette, Andre J.
TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
TITLE OF INVENTION: of Their Use
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California

; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/342,268
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 1206
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-342-268-36

Query Match 13.3%; Score 6; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 13
US-09-015-968-33
; Sequence 33, Application US/09015968
; Patent No. 6057425
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,279
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 1206
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949

; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 3003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-015-968-33

Query Match 13.3%; Score 6; DB 3; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 LVLLAF 22
Db 2 LVLLAF 7

RESULT 14
US-09-015-968-36
; Sequence 36, Application US/09015968
; Patent No. 6057425
; GENERAL INFORMATION:
; APPLICANT: Selsted, Michael E.
; APPLICANT: Ouellette, Andre J.
; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods
; TITLE OF INVENTION: of Their Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,968
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,279
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/342,268
; FILING DATE: 18-NOV-1994
; APPLICATION NUMBER: US 07/930,649
; FILING DATE: 14-AUG-1992
; APPLICATION NUMBER: US 07/889,020
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-UC 3003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

US-09-015-968-36

Query Match 13.3%; Score 6; DB 3; Length 85;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22

Db 2 LVLLAF 7

RESULT 15

US-09-397-386-33

; Sequence 33, Application US/09397386

; Patent No. 6300470

; GENERAL INFORMATION:

; APPLICANT: Selsted, Michael E.

; APPLICANT: Ouellette, Andre J.

; TITLE OF INVENTION: Antibiotic Cryptdin Peptides and Methods

; TITLE OF INVENTION: of Their Use

; NUMBER OF SEQUENCES: 70

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/397,386

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/015,968

; FILING DATE:

; APPLICATION NUMBER: US 08/482,279

; FILING DATE: 07-JUN-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/342,268

; FILING DATE: 18-NOV-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/930,649

; FILING DATE: 14-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/889,020

; FILING DATE: 26-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-UC 3003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 33:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 85 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-09-397-386-33

Query Match

13.3%; Score 6; DB 4; Length 85;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 LVLLAF 22

Db 2 LVLLAF 7

Search completed: March 4, 2003, 17:02:24

Job time : 45.5424 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)


```

; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745.008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
; US-09-745-008-34

Query Match      62.2%; Score 28; DB 10; Length 1162;
Best Local Similarity 100.0%; Pred. No. 3.1e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPLRRQVVVVPLSPRLVLLAFRCRQLP 28
Db 356 QPLRRQVVVVPLSPRLVLLAFRCRQLP 383

RESULT 3
US-09-745-008-13
; Sequence 13, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745.008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-13

Query Match      33.3%; Score 15; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 KRMGGSYRCVNASTA 44
Db 6 KRMGGSYRCVNASTA 20

RESULT 4
US-09-745-008-15
; Sequence 15, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745.008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 21

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-15

Query Match      22.2%; Score 10; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 YRCVNASTAN 45
Db 12 YRCVNASTAN 21

RESULT 5
US-09-955-909-4
; Sequence 4, Application US/09955909
; Patent No. US20020150995A1
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: SIALYLLOIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/955,909
; FILING DATE: 18-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,393
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 642 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-955-909-4

Query Match      22.2%; Score 10; DB 10; Length 642;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 YRCVNASTAN 45
Db 437 YRCVNASTAN 446

```



```
RESULT 6
US-09-745-008-2
; Sequence 2, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-2

Query Match      22.2%; Score 10; DB 10; Length 666;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
   |||||
Db 469 YRCVNASTAN 478

RESULT 7
US-09-955-909-2
; Sequence 2, Application US/09955909
; Patent No. US20020150995A1
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; BARKER, William A.
; HAKES, David J.
; ZOFF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; SIALYLLOIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/955,909
; FILING DATE: 18-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,393
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Cortuzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-955-909-2

Query Match      22.2%; Score 10; DB 10; Length 1060;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 YRCVNASTAN 45
   |||||
Db 468 YRCVNASTAN 477

RESULT 8
US-09-745-008-14
; Sequence 14, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-14

Query Match      20.0%; Score 9; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 KRMGGSYRC 38
   |||||
Db 6 KRMGGSYRC 14

RESULT 9
US-09-745-008-17
; Sequence 17, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-17
```

Query Match 15.6%; Score 7; DB 10; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 VNASTAN 45
Db 1 VNASTAN 7

RESULT 10

US-10-084-018-8
; Sequence 8, Application US/10084018
; Patent No. US20020160499A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; Hawkins, Phillip R.
; Hillman, Jennifer L.
; Lal, Pretti
; Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN SERINE
; CARBOXYPEPTIDASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,018
FILING DATE: 25-Feb-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/299,689A

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/828,488

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PP-0241 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 480 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 190283

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-10-084-018-8

Query Match 15.6%; Score 7; DB 9; Length 480;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 RLPLKRM 32
Db 307 RLPLKRM 313

RESULT 11

US-09-736-457-336
; Sequence 336, Application US/09736457

Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 336
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-336

Query Match 15.6%; Score 7; DB 9; Length 480;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 RLPLKRM 32
Db 307 RLPLKRM 313

RESULT 12

US-09-902-941-336
; Sequence 336, Application US/09902941
; Patent No. US20020172952A1

GENERAL INFORMATION:

APPLICANT: Henderson, Robert A.

APPLICANT: Wang, Tongtong

APPLICANT: Watanabe, Yoshihiro

APPLICANT: Johnson, Jeffrey C.

APPLICANT: Retter, Marc W.

APPLICANT: Marnerakis, Margarita

APPLICANT: Carter, Darrick

APPLICANT: Fanger, Gary R.

APPLICANT: Vedvick, Thomas S.

APPLICANT: Bangur, Chaitanya S.

APPLICANT: McNabb, Andria

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.478C17

CURRENT APPLICATION NUMBER: US/09/902,941

CURRENT FILING DATE: 2001-07-10

NUMBER OF SEQ ID NOS: 2002

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 336

LENGTH: 480

TYPE: PRT

ORGANISM: Homo sapiens

US-09-902-941-336

Query Match 15.6%; Score 7; DB 9; Length 480;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 RLPLKRM 32
Db 307 RLPLKRM 313

RESULT 13

US-09-849-626-336
; Sequence 336, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Ranger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 336
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-336

Query Match 15.6%; Score 7; DB 9; Length 480;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RLPLKRM 32
|||||
DB 307 RLPLKRM 313

RESULT 14
US-09-925-302-774
; Sequence 774, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 774
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-774

Query Match 15.6%; Score 7; DB 10; Length 492;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RLPLKRM 32
|||||
DB 319 RLPLKRM 325

RESULT 15
US-09-815-242-11760
; Sequence 11760, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11760
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11760

Query Match 15.6%; Score 7; DB 10; Length 577;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RVVVVPL 13
|||||
DB 402 RVVVVPL 408

Search completed: March 4, 2003, 17:02:57
Job time : 15.4915 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:50:35 ; Search time 14.9492 Seconds
(without alignments) 124.790 Million cell updates

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Scoring table: 2500002
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 2000000000

post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum Match 100%
Listing first 45 summaries

Database : A Genesec 101002:*

1: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID\$2/cgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	53.8	30	AAW9049	Polypeptide fragme
2	42	53.8	30	ABB51214	Human secreted pro
3	42	53.8	339	ABB68550	Drosophila melanog
4	41	52.6	189	AAU2341	Propionibacterium
5	41	52.6	212	ABB75629	Human endo type pr
6	40.5	51.9	235	AAK05328	Rat tumour necrosi
7	40.5	51.9	235	AAK05996	Rat tumour necrosi
8	40.5	51.9	235	AAU10546	Rat tumour necrosi
9	40	51.3	92	AAAG3234	Zea mays protein f
10	40	51.3	138	AAAG3233	Zea mays protein f

11	40	51.3	151	22	ABG16755	Novel human diago
12	40	51.3	230	22	ABG98324	Escherichia coli p
13	40	51.3	299	22	ABG76024	Human protein with
14	40	51.3	428	23	ABB55482	Lactococcus lactis
15	40	51.3	483	21	AAG34982	Zea mays protein f
16	40	51.3	529	21	AAG34981	Zea mays protein f
17	40	51.3	559	23	ABP43486	Human secreted pro
18	40	51.3	565	21	AGA43980	Zea mays protein f
19	40	51.3	570	22	ABB76023	Neurotrophin-like p
20	40	51.3	586	22	ABB76018	Neurotrophin-like p
21	40	51.3	1219	22	ABG12100	Novel human diago
22	39.5	50.6	732	20	AAI42697	Mouse serine-threo
23	39.5	50.6	733	20	AAI42696	Rat serine-threoni
24	39.5	50.6	806	22	ABG56522	Novel protein kina
25	39	50.0	71	22	AAU40677	Propionibacterium
26	39	50.0	71	22	AAU66518	Propionibacterium
27	39	50.0	126	22	AAU24153	Human EST encoded
28	39	50.0	191	22	ABG01352	Novel human diago
29	39	50.0	356	22	AAU02417	Human novel melast
30	39	50.0	366	22	AAU02411	Human novel melast
31	39	50.0	376	22	AAU02418	Human novel melast
32	39	50.0	386	22	AAU02412	Human novel melast
33	39	50.0	429	22	ABG07361	Novel human diago
34	39	50.0	501	22	ABU02419	Human novel melast
35	39	50.0	511	22	AAU02413	Human novel melast
36	39	50.0	542	17	AAR99389	Mouse 030 gene (fo
37	39	50.0	542	18	AAW26668	Mouse tumour progr
38	39	50.0	542	21	AAI76915	Mouse fomy030 prot
39	39	50.0	542	22	AAE05106	Murine homologue,
40	39	50.0	542	23	AAU96920	Mouse fomy 030 pro
41	39	50.0	542	23	AAW668130	Mouse fomy030 poly
42	39	50.0	542	23	AAE13593	Mouse fomy 030 pro
43	39	50.0	687	22	AAU02420	Human novel melast
44	39	50.0	697	22	AAU02414	Human novel melast
45	39	50.0	750	22	AAU02421	Human novel melast

ALIGNMENTS

RESULT 1	
AAW89049	Human; secreted protein; fusion protein; diagnosis; tissue; cancer; tumour; neurodegenerative disease; Alzheimer's disease; developmental abnormality; foetal deficiency; immune system; asthma; lymphocytic diseases; inflammation; ischaemic shock; Alzheimer's cognitive disorder; schizophrenia; prostatic osteoporosis; arthritis; testis; lung; thyroid; endocrine; metabolism; regulation; malabsorption
AAW89049	Homo sapiens.
AAW89049	WO9854963-A2.
01-MAR-1999	10-DEC-1998.
(first entry)	04-JUN-1998; 98WO-US11422.
Polypeptide fragment encoded by gene 183.	18-DEC-1997; 97US-0070923.
Human; secreted protein; fusion protein; diagnosis; tissue; cancer; tumour; neurodegenerative disease; Alzheimer's disease; developmental abnormality; foetal deficiency; immune system; asthma; lymphocytic diseases; inflammation; ischaemic shock; Alzheimer's cognitive disorder; schizophrenia; prostatic osteoporosis; arthritis; testis; lung; thyroid; endocrine; metabolism; regulation; malabsorption	06-JUN-1997; 97US-0048877.
Homo sapiens.	06-JUN-1997; 97US-0048881.
WO9854963-A2.	06-JUN-1997; 97US-0048884.
10-DEC-1998.	06-JUN-1997; 97US-0048893.
04-JUN-1998; 98WO-US11422.	06-JUN-1997; 97US-0048896.

PR 06-JUN-1997; 97US-0048899.
 PR 06-JUN-1997; 97US-0048915.
 PR 06-JUN-1997; 97US-0048949.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048972.
 PR 06-JUN-1997; 97US-0049020.
 PR 06-JUN-1997; 97US-0049375.
 PR 05-SEP-1997; 97US-0057628.
 PR 05-SEP-1997; 97US-0057635.
 PR 05-SEP-1997; 97US-0057644.
 PR 05-SEP-1997; 97US-0057647.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057661.
 PR 05-SEP-1997; 97US-0057667.
 PR 05-SEP-1997; 97US-0057761.
 PR 05-SEP-1997; 97US-0057764.
 PR 05-SEP-1997; 97US-0057770.
 PR 05-SEP-1997; 97US-0057775.
 PR 05-SEP-1997; 97US-0057778.
 PR 06-JUN-1997; 97US-0048875.
 PR 06-JUN-1997; 97US-0048878.
 PR 06-JUN-1997; 97US-0048882.
 PR 06-JUN-1997; 97US-0048885.
 PR 06-JUN-1997; 97US-0048894.
 PR 06-JUN-1997; 97US-0048897.
 PR 06-JUN-1997; 97US-0048900.
 PR 06-JUN-1997; 97US-0048916.
 PR 06-JUN-1997; 97US-0048962.
 PR 06-JUN-1997; 97US-0048970.
 PR 06-JUN-1997; 97US-0048974.
 PR 06-JUN-1997; 97US-0049373.
 PR 05-SEP-1997; 97US-0057584.
 PR 05-SEP-1997; 97US-0057629.
 PR 05-SEP-1997; 97US-0057642.
 PR 05-SEP-1997; 97US-0057645.
 PR 05-SEP-1997; 97US-0057648.
 PR 05-SEP-1997; 97US-0057651.
 PR 05-SEP-1997; 97US-0057662.
 PR 05-SEP-1997; 97US-0057668.
 PR 05-SEP-1997; 97US-0057762.
 PR 05-SEP-1997; 97US-0057765.
 PR 05-SEP-1997; 97US-0057771.
 PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 05-SEP-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057666.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;

PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX WPI; 1999-059865/05.
 DR N-PSDB; AAV84593.
 XX New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX Disclosure; Page 143; 772pp; English.
 XX The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
 CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents a polypeptide fragment encoded by a
 CC gene of the invention (see descriptor line for gene number).
 XX SQ Sequence 30 AA;
 Query Match 53.8%; Score 42; DB 20; Length 30;
 Best Local Similarity 61.5%; Pred. No. 2.7;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 RQRLPKRMGGSYR 13
 DB 5 RQRRPRRGTSR 17
 RESULT 2
 ABB51214
 ID ABB51214 standard; Protein; 30 AA.
 XX ABB51214;
 AC ABB51214;
 XX 07-FEB-2002 (first entry)
 DE Human secreted protein encoded by gene 183 SEQ ID NO:1167.
 XX Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;
 KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 KW cyostatic; cardiac; vascular; anti-angiogenic; ophthalmologic;
 KW neuroprotective; nootropic; anticonvulsant; antialzheimers; vulnary;
 KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
 KW human immunodeficiency virus; hyperproliferative disorder; wound healing;
 KW Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;
 KW Chaga's cardiomyopathy; coronary arteriosclerosis; angiodenic disorder;
 KW corneal graft neovascularisation; diabetic retinopathy; regeneration;
 KW neurological disorder; Huntington's chorea; Alzheimer's disease;
 KW Parkinson's disease; infectious disease; chromosome 17.
 XX Homo sapiens.

XX WO200162891-A2.
 PN XX
 PD XX
 PP 30-AUG-2001.
 XX
 XX 21-FEB-2001; 2001WO-US05614.
 XX
 XX 24-FEB-2000; 2000US-184836P.
 PR 29-MAR-2000; 2000US-193170P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Ni J, Ebner R, Lafleur DW, Moore PA, Olsen HS, Rosen CA;
 PI Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;
 PI Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrie AM, Fan P;
 PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;
 PI Zeng Z, Greene JM;
 XX WPI; 2001-625724/72.
 DR
 XX Nucleic acids encoding 207 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 PT
 XX Disclosure; Page 398; 1533pp; English.
 PS
 XX ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted
 CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various
 CC activities based on the tissues and cells the genes are expressed in.
 CC Example of these activities include: immunomodulatory; anticlerotic;
 CC dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 CC anti-HIV; cytostatic; cardiant; anti-angiogenic; ophthalmological;
 CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;
 CC antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used
 CC in gene therapy and vaccine production. (I) and (II) can be used in the
 CC prevention, diagnosis and treatment of immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to
 CC ABA83193 and ABB50300 represent sequences used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 30 AA;
 Query Match 53.8%; Score 42; DB 22; Length 30;
 Best Local Similarity 61.5%; Pred. No. 2.7;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 OY 1 RQRLPKRMGGSYR 13
 |||:|:|:|
 Db 5 RQRRPRRGCGTSR 17
 RESULT 3
 ABB68550
 ID ABB68550 standard; Protein; 339 AA.
 XX
 AC ABB68550;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 32442.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL12653.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 32442; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 339 AA;
 Query Match 53.8%; Score 42; DB 22; Length 339;
 Best Local Similarity 70.0%; Pred. No. 33;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 5 PKEMGGSYRC 14
 |||:|:|:|
 Db 208 PKRVPGPYRC 217
 RESULT 4
 AAU42341
 ID AAU42341 standard; Protein; 189 AA.
 XX
 AC AAU42341;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #3237.
 XX
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 XX WO200181581-A2.
 PN
 XX 01-NOV-2001.
 PD
 XX 20-APR-2001; 2001WO-US12865.
 PF
 XX 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI; 2001-616774/71.
 DR N-PSDB; AAS59516.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 XX Example 1; SEQ ID No 3536; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 189 AA;

Query Match 52.6%; Score 41; DB 22; Length 189;
 Best Local Similarity 61.5%; Pred. No. 27;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYR 13
 |||||: |||:
 DB 177 RQRLPPQAGSHR 189

RESULT 5
 ABB75629
 ID ABB75629 standard; Protein; 212 AA.
 XX
 AC ABB75629;
 XX
 DT 10-JUN-2002 (first entry)
 XX
 DE Human endo type protease 23.32.
 XX
 KW Endo type protease 23.32; endoprotease; human; tumour; haemopathy;
 KW HIV infection; immunological disease; inflammation; cytostatic;
 KW haemostatic; anti-HIV; virucide; immunomodulator; antiinflammatory;
 KW enzyme; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200220744-A1.
 XX
 PD 14-MAR-2002.
 XX
 PF 02-JUL-2001; 2001WO-CN01144.
 XX
 PR 07-JUL-2000; 2000CN-0119412.
 XX
 PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
 XX
 PI Mao Y, Xie Y;

XX WPI; 2002-369623/31.
 DR N-PSDB; ABL53561.
 XX
 PT Human endo type protease 23.32 and encoding polynucleotide, used in
 PT diagnosis and treatment of malignant tumors, hemopathy, human
 PT immunodeficiency virus infection, immunological diseases and
 PT inflammation -
 XX
 PS Claim 1; Page 30; 36pp; Chinese.
 XX
 CC The present sequence is the protein sequence for human endo type
 CC protease 23.32. The protease protein and its coding sequence are
 CC useful for the diagnosis and treatment of malignant tumors,
 CC haemopathy, HIV infection, immunological disease and inflammation.
 XX
 SQ Sequence 212 AA;

Query Match 52.6%; Score 41; DB 23; Length 212;
 Best Local Similarity 63.6%; Pred. No. 31;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGSYR 13
 |||||: |||:
 DB 198 RLPKRMGGENR 208

RESULT 6
 AAR05328
 ID AAR05328 standard; protein; 235 AA.
 XX
 AC AAR05328;
 XX
 DT 03-DEC-1990 (first entry)
 XX
 DE Rat tumour necrosis factor (TNF) gene product.
 XX
 KW Tumour necrosis factor; TNF; antioncotic; cancer; ds.
 XX
 OS Rattus sp.
 XX
 PN JP02157295-A.
 XX
 PD 18-JUN-1990.
 XX
 PF 07-DEC-1988; 88JP-0307751.
 XX
 PR 07-DEC-1988; 88JP-0307751.
 XX
 PA (ASAH) ASAH CHEMICAL IND KK.
 XX
 DR WPI; 1990-228715/30.
 DR N-PSDB; AAR05328.
 XX
 PT Novel anti-oncotic polypeptide - prepd. by culturing microbe
 PT contg. DNA.
 XX
 PS Disclosure; ; Japanese.
 XX
 CC Fragment of gene product (AAs 80-235) is claimed as an antioncotic
 CC peptide, produced from a transformed microorganism expression
 CC system.
 XX
 SQ Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 11; Length 235;
 Best Local Similarity 56.2%; Pred. No. 42;
 Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14
 : ||||: |||:
 DB 15 EALPKRMGGGLQNSRR 30

RESULT 7
AAU05996
ID AAR05996 standard; protein; 235 AA.

XX AAR05996;
AC AAR05996;

DT 03-DEC-1990 (first entry)

DE Rat tumour necrosis factor (TNF).

KW Tumour necrosis factor; TNF; antioncotic; cancer; ds.

OS Rattus sp.

PN JP02157295-A.

PD 18-JUN-1990.

XX 07-DEC-1988; 88JP-0307751.

XX 07-DEC-1988; 88JP-0307751.

XX (ASAHI) ASAHI CHEMICAL IND KK.

XX WPI; 1990-228715/30.

DR P-PSDB; AAR05996.

XX Novel anti-oncotic polypeptide - prepd. by culturing microbe contg. DNA.

PS Disclosure; ; Japanese.

XX Fragment of gene product (AAS 80-235) is claimed as an antioncotic peptide, produced from a transformed microorganism expression system.

XX Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 11; Length 235;
Best Local Similarity 56.2%; Pred. No. 42;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

OY 2 QRLPKRMGG---SYRC 14

DB 15 EALPKKMGGLQNSRRC 30

RESULT 8
AAU10546
ID AAU10546 standard; Protein; 235 AA.

XX AAU10546;

DT 14-FEB-2002 (first entry)

DE Rat tumour necrosis factor (TNF) alpha (YSG10) polypeptide.

XX YSG; YSG10; schizophrenia; chronic animal model; LCGU; netrin receptor;
KW local cerebral glucose utilisation; phosphodiesterase 1-alpha; UNC5H1;
KW calcium-independent alpha-latrotoxin receptor; CIRL; trkE; synapsin 1A;
KW epithelial discoidin domain receptor 1; synapsin 1B; neuroleptic; ss;
KW tumour necrosis factor alpha; TNF-alpha; rat.

XX Rattus sp.

XX Key Location/Qualifiers

FT Misc-difference 190

FT /note= "Encoded by GAG"

XX WO200175440-A2.

XX 11-OCT-2001.

XX 02-APR-2001; 2001WO-GB01486.

XX 31-MAR-2000; 2000GB-0007880.

XX 26-MAY-2000; 2000GB-0012768.

XX (WELF-) WELFIDE CORP.

XX Cochran S.; Paterson G.; Ohashi Y.; Morris B.; Pratt J;

XX WPI; 2002-010813/01.

XX N-PSDB; AAS16846.

XX Novel chronic animal model of schizophrenia, useful for identifying

XX anti-psychotic drugs and genes that are associated with schizophrenia

XX Disclosure; Fig 10b; 79pp; English.

XX The invention relates to YSG polynucleotide fragments for use in
XX diagnosing and/or developing treatments for schizophrenia using chronic
XX animal models. The polynucleotides and their encoded polypeptides are
XX used for identification of compounds which modulate the expression of YSG
XX molecules, leading to the manufacture of schizophrenia medicaments. The
XX sequences can also be used for testing candidate compounds for any effect
XX on the polypeptides. Anti-schizophrenic effects of a compound can be
XX determined by measuring local cerebral glucose utilisation (LCGU) or
XX comparing its expression level with that of a control group. The
XX sequences are useful in the identification of genes associated with
XX schizophrenic states and in the development of an antibody. The sequences
XX of the invention include phosphodiesterase 1-alpha, calcium-independent
XX alpha-latrotoxin receptors (CIRL)-1,2k3, epithelial discoidin domain
XX receptor 1 (trkE), netrin receptor (UNC5H1), synapsin 1A and AB and
XX tumour necrosis factor (TNF) alpha. This sequence represents rat tumour
XX necrosis factor (TNF) alpha (YSG10) polypeptide.

XX Sequence 235 AA;

Query Match 51.9%; Score 40.5; DB 23; Length 235;
Best Local Similarity 56.2%; Pred. No. 42;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

OY 2 QRLPKRMGG---SYRC 14

DB 15 EALPKKMGGLQNSRRC 30

RESULT 9
AAG33234
ID AAG33234 standard; Protein; 92 AA.

XX AAG33234;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 40237.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.

OS Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.

```
PR 08-OCT-1999; 99US-01582322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.3%; Score 40; DB 21; Length 92;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 RLPKRMGGSY 12
Db 47 RLPVAVGGSY 56

RESULT 10
AAG33233
ID AAG33233 standard; Protein; 138 AA.
XX AC AAG33233;
XX DT DT (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 40236.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
```

PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145192.
 PR 22-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154039.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 51.3%; Score 40; DB 21; Length 138;
 Best Local Similarity 80.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 RLPKRMGGSY 12
 |||||
 DB 93 RLPVAVGGSY 102

RESULT 11

ABGL6755
 ID ABGL6755 standard; Protein; 151 AA.

XX AC ABGL6755;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #16746.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS80942.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity -

XX XX Claim 20; SEQ ID No 47114; 103pp; English.

XX PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 151 AA;
 Query Match 51.3%; Score 40; DB 22; Length 151;
 Best Local Similarity 63.6%; Pred. No. 32;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 11
 | | | | |
 Db 78 RPRAPKRLGGA 88

RESULT 12
 AAG98324
 ID AAG98324 standard; Protein; 230 AA.

XX AC AAG98324;
 XX DT 21-SEP-2001 (first entry)
 XX DE Escherichia coli protein sequence SEQ ID NO:372.
 XX KW Escherichia coli; identification; proliferation; microorganism;
 XX KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 XX KW bacterial growth inhibition.
 XX OS Escherichia coli.
 XX PN WO200148209-A2.

XX PD 05-JUL-2001.
 XX PF 19-DEC-2000; 2000WO-US34419.
 XX PR 23-DEC-1999; 99US-0173005.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Forsyth RA, Ohlsen KL, Zyskind JW;
 XX DR WPI; 2001-457376/49.
 XX DR N-PSDB; AAH81380.

XX PT Novel nucleic acids encoding proteins required for Escherichia coli
 XX proliferation, useful for screening for antimicrobial agents -
 XX PS Claim 19; Page 496-497; 596pp; English.

XX The present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (I) in a microorganism inhibits proliferation

CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.
 CC For example, nucleic acid probes complementary to proliferation-required
 CC sequences that are specific for particular species of microorganisms can
 CC be used as probes to identify particular microorganism species in
 CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
 CC proteins given in AAG98324 to AAG98431, and AAH81488 to AAH81491
 CC represent oligonucleotides, which are used in the exemplification of the
 CC present invention.

SQ Sequence 230 AA;

Query Match 51.3%; Score 40; DB 22; Length 230;
 Best Local Similarity 70.0%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 10
 | | | | |
 Db 207 KQRIKRMGG 216

RESULT 13
 ABB76024
 ID ABB76024 standard; Protein; 299 AA.

XX AC ABB76024;
 XX DT 12-JUL-2002 (first entry)
 XX DE Human protein with immunoglobulin domain.
 XX KW Neurotrophin-like polypeptide; antiparkinsonian; neurotropic;
 XX KW anticonvulsant; neuroprotective; cerebroprotective; antidepressant;
 XX KW tranquilizer; hypotensive; gene therapy; immunoglobulin; human.
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT Domain 1..200
 XX FT /note= "immunoglobulin domain"
 XX PN WO200157175-A2.

XX PD 09-AUG-2001.
 XX PF 02-FEB-2001; 2001WO-US03651.
 XX PR 03-FEB-2000; 2000US-0496914.
 XX PR 27-APR-2000; 2000US-0560875.
 XX PR 02-AUG-2000; 2000US-0632085.
 XX PA (HYSE-) HYSEQ INC.

XX PI Boyle BJ, Mize NK, Arterburn MC, Tang YT, Yeung G, Zhou P, Liu C;
 XX PI Asundi V, Drmanac RT, Wang M, Chen L, Yang Y;
 XX DR WPI; 2001-476282/51.

XX PT Novel neurotrophin-like polypeptide is useful in treating
 XX PT neurodegenerative, immune, etc. disorders -
 XX PS Disclosure; Page 131-132; 134pp; English.

XX The present sequence is that of a novel human protein with an
 CC immunoglobulin domain. Its sequence shows 100% similarity over
 CC 200 amino acid residues to novel human neurotrophin-like polypeptide
 CC (see ABB76017), and 99% similarity over 299 amino acid residues to
 CC the splice variant given in ABB76018. The invention provides

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145986.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157553.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.3%; Score 40; DB 21; Length 483;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 RLPKMGGSY 12
DB 47 RLPVNGGSY 56

Search completed: March 4, 2003, 16:53:18
Job time : 15.9492 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:51:40 ; Search time 4.27119 Seconds
(without alignments)
135.950 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	78	100.0	1162	1 TCNA_TRYCR	P23253 trypanosoma
2	40.5	51.9	235	1 TNFA_RAT	P16599 rattus norv
3	40	51.3	210	1 YQEH_ECOLI	Q46941 escherichia
4	40	51.3	428	1 YL28_LACLA	Q9cdt3 lactococcus
5	39	50.0	304	1 K2S3_HUMAN	Q14952 homo sapien
6	39	50.0	311	1 STCC_EMENI	Q00668 emericeila
7	39	50.0	520	1 IBMP_CAVPO	P18617 cauliflower
8	39	50.0	1319	1 MN1_HUMAN	Q10571 homo sapien
9	38	48.7	361	1 IDH1_KLULA	O94229 kluyveromyc
10	38	48.7	553	1 GKP2_HUMAN	Q14410 homo sapien
11	38	48.7	637	1 SCAA_CHICK	Q92075 gallus gall
12	38	48.7	1403	1 PRO_DROME	P29617 drosophila
13	38	48.7	1556	1 RS1_DROVI	Q9661 drosophila
14	37.5	48.1	54	1 RL14_AERPE	P58731 aeropyrum p
15	37	47.4	197	1 IE68_HSV2	P14379 herpes simp
16	37	47.4	253	1 YDEO_ECOLI	P76135 escherichia
17	37	47.4	304	1 K2S1_HUMAN	Q14954 homo sapien
18	37	47.4	304	1 K2S2_HUMAN	P43631 homo sapien
19	37	47.4	304	1 K2S4_HUMAN	P43632 homo sapien
20	37	47.4	304	1 K2S5_HUMAN	Q14953 homo sapien
21	37	47.4	341	1 K2L3_HUMAN	P43628 h killer ce
22	37	47.4	348	1 K2L1_HUMAN	P43626 h killer ce
23	37	47.4	348	1 K2L2_HUMAN	P43627 homo sapien
24	37	47.4	387	1 K3S1_HUMAN	Q14943 homo sapien
25	37	47.4	444	1 K3L1_HUMAN	P43629 h killer ce
26	37	47.4	455	1 K3L2_HUMAN	P43630 homo sapien
27	37	47.4	504	1 MPFA_SOLTU	P29677 solanum tub
28	37	47.4	520	1 IBMP_CAVPO	P03559 cauliflower
29	37	47.4	646	1 MU18_HUMAN	P43121 homo sapien
30	36.5	46.8	234	1 TNFA_CAVPO	P51435 cavia porce
31	36.5	46.8	235	1 TNFA_MOUSE	P06804 mus musculu
32	36	46.2	87	1 RL34_SULTO	Q975K6 sulfolobus
33	36	46.2	89	1 RL34_METJA	P54053 methanococc

34	36	46.2	97	1 SY07_MOUSE	Q03366 mus musculu
35	36	46.2	97	1 SY07_RAT	Q9gxy8 rattus norv
36	36	46.2	192	1 VMTZ_LAMB	P03731 bacterioph
37	36	46.2	196	1 RL11_CAEL	Q94300 caenorhabdi
38	36	46.2	221	1 FIXW_RHILE	P14312 rhizobium l
39	36	46.2	255	1 YORE_PSECL	Q03003 pseudomonas
40	36	46.2	452	1 MTN3_CHICK	Q42401 gallus gall
41	36	46.2	473	1 FP2_MYTGA	Q25464 mytilus gal
42	36	46.2	497	1 SC14_YARLI	P45816 yarrowia li
43	36	46.2	520	1 IBMP_CAVPO	P22547 cauliflower
44	36	46.2	828	1 MRKC_KLEPN	P21647 klebsiella
45	36	46.2	2215	1 MY7A_HUMAN	Q13402 homo sapien

ALIGNMENTS

RESULT 1
TCNA_TRYCR STANDARD; PRT: 1162 AA.
AC P23253;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen).
GN TCNA.

OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Silvio X-10/4;
RX MEDLINE=91277609; PubMed=1711561;
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,
RA Prioli R.P.;
RT "The Trypanosoma cruzi neuraminidase contains sequences similar to
bacterial neuraminidases, YWTD repeats of the low density lipoprotein
receptor, and type III modules of fibronectin."
RL J. Exp. Med. 174:179-191(1991).
RN [2]

RP SUBCELLULAR LOCATION.
RX MEDLINE=91376547; PubMed=1896773;
RA Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;
RT "Trypanosoma cruzi: localization of neuraminidase on the surface of
trypanastigotes."
RL Trop. Med. Parasitol. 42:146-150(1991).
CC -!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN
PARASITIC INVASION OF CELLS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
alpha-(2->8)-glycosidic linkages of terminal sialic residues in
oligosaccharides, glycoproteins, glycolipids, colominic acid and
synthetic substrates.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POSSIBLE).
CC -!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM
IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.
CC -!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT
DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TNCA
PROTEIN.

CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.

CC -!- SIMILARITY: CONTAINS 3 ENR REPEATS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC -----

DR EMBL; M61732; AAA30255.1; --

DR PIR; JH0557; JH0557.

DR InterPro; IPR002860; GH_BNR.

DR Pfam; PF02012; BNR; 2.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 100.0%; Score 78; DB 1; Length 1162;
Best Local Similarity 100.0%; Pred. No. 4.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14
|||||
Db 379 RQRLPKRMGGSYRC 392

RESULT 2
TNFA RAT
ID TNFA RAT STANDARD; PRT; 235 AA.
AC P16599; 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor precursor (TNF-alpha) (Tumor necrosis factor
ligand superfamily member 2) (Cachectin).
GN TNF OR TNFSF2 OR TNFA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=94040766; PubMed=8224868;
RA Kwon J., Chung I.Y., Benveniste E.N.;
RT "Cloning and sequence analysis of the rat tumor necrosis
factor-encoding genes";
RL Gene 132:227-236 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Shirai T., Shimizu N., Horiguchi S., Ito H.;
RT "Cloning and expression in Escherichia coli of the gene for rat tumor
necrosis factor";
RL Agric. Biol. Chem. 53:1733-1736 (1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92329007; PubMed=1627266;
RA Batler H.C., Grewe M., Gaussling R., Pavlovic M., Decker K.;
RT "Rat tumor necrosis factor-alpha. Transcription in rat Kupffer cells
and in vitro posttranslational processing based on a PCR-derived
cDNA";
RL Biol. Chem. Hoppe-Seyler 373:271-281 (1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Various;
RX MEDLINE=21369712; PubMed=11477479;
RA Furuya T., Joe B., Salstrom J.L., Hashimoto A., Dobbins D.E.,
RA Wilder R.L., Remmers E.F.;
RT "Polymorphisms of the tumor necrosis factor alpha locus among
autoimmune disease susceptible and resistant inbred rat strains.";
RN Genes Immun. 2:229-232 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RA Decker K.F.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [6]

RP Pfam; PF02012; BNR; 2.
RC Kirisits M.J., Vardimon D., Kunz H.W., Gill T.J. III;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cytokine that binds to TNFRSF1A/TNFR1 and
CC TNFRSF1B/TNFR. It is mainly secreted by macrophages and can
CC induce cell death of certain tumor cell lines. It is potent
CC pyrogen causing fever by direct action or by stimulation of
CC interleukin 1 secretion and is implicated in the induction of
CC cachexia. Under certain conditions it can stimulate cell
CC proliferation and induce cell differentiation.
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Also exists as an
CC extracellular soluble form (By similarity).
CC -!- PTM: The soluble form derives from the membrane form by
CC proteolytic processing (By similarity).
CC -!- PTM: The membrane form, but not the soluble form, is
CC phosphorylated on serine residues. Dephosphorylation of the
CC membrane form occurs by binding to soluble TNFRSF1A/TNFR1 (By
CC similarity).
CC -!- DISEASE: CACHEXIA ACCOMPANIES A VARIETY OF DISEASES, INCLUDING
CC CANCER AND INFECTION, AND IS CHARACTERIZED BY GENERAL ILL HEALTH
CC AND MALNUTRITION.
CC -!- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X66539; CAA47146.1; -.
DR ENBL; L00981; AAA16275.1; -.
DR ENBL; D00475; BAA00367.1; -.
DR ENBL; AJ002278; CAA05290.1; -.
DR ENBL; L19123; AAA42255.1; -.
DR ENBL; AF329982; AAK53568.1; -.
DR ENBL; AF329983; AAK53569.1; -.
DR ENBL; AF329984; AAK53570.1; -.
DR ENBL; AF329985; AAK53571.1; -.
DR ENBL; AF329986; AAK53572.1; -.
DR ENBL; AF329987; AAK53573.1; -.
DR PIR; JU0029; JU0029.
DR PIR; S21674; S21674.
DR PIR; JN0868; JN0868.
DR HSP; P06804; 2TNF.
DR InterPro; IPR003636; TNF abc.
DR InterPro; IPR000478; TNF_family.
DR Pfam; PF00229; TNF; 1.
DR PRINTS; PR01234; TNECROSISFCT.
DR ProDom; PD002012; TNF abc; 1.
DR SMART; SM00207; TNF; 1.
DR PROSITE; PS00251; TNF 1; 1.
DR PROSITE; PS00049; TNF 2; 1.
KW Cytokine; Cytocoxin; Transmembrane; Glycoprotein; Phosphorylation;
KW Signal-anchor.
FT CHAIN 1 235 TUMOR NECROSIS FACTOR, MEMBRANE FORM.
FT CHAIN 80 235 TUMOR NECROSIS FACTOR, SOLUBLE FORM.
FT DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 36 56 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 57 235 EXTRACELLULAR (POTENTIAL).
FT SITE 79 80 CLEAVAGE (BY ADAM17) (BY SIMILARITY).
FT MOD RES 2 2 PHOSPHORYLATION (BY CK1) (BY SIMILARITY).
FT DISULFID 148 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 86 86 L -> P (IN REF. 3 AND 5).
FT CONFLICT 39 39 L -> T (IN REF. 3 AND 5).
FT CONFLICT 163 163 F -> S (IN REF. 3 AND 5).
FT CONFLICT 202 202
SQ SEQUENCE 235 AA; 25806 MW; B808EC6D049C2F3B CRC64;

```

Query Match          51.3%; Score 40.5; DB 1; Length 235;
Best Local Similarity 56.2%; Pred. No. 5;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14
Db 15 EALPKMGGLQNSRR 30

RESULT 3
YOEK ECOLI
ID YOEK_ECOLI STANDARD; PRT; 210 AA.
AC Q46941;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yqeh.
GN YOEK OR B2846.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SIMILARITY: SOME, TO E.COLI YKKG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29581; AAB40493.1; ALT_INIT.
DR EMBL; AE000368; AAC75885.1; ALT_INIT.
DR EcoGene; EGI30399; Yqeh.
DR InterPro; IPR000792; HTH LuxR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 210 AA; 24287 MW; 9C07787FEC9823C CRC64;

Query Match          51.3%; Score 40; DB 1; Length 210;
Best Local Similarity 70.0%; Pred. No. 5.5;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRLPKRMGG 10
Db 187 QRIEKMGG 196

RESULT 4
YL28 LACLA
ID YL28_LACLA STANDARD; PRT; 428 AA.
AC Q9CDDT3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc metalloprotease LL2128 (EC 3.4.24.-).
GN LL2128.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;

Query Match          51.3%; Score 40; DB 1; Length 428;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PKRMGGSYR 13
Db 280 PKRMGGSYR 288

RESULT 5
K2S3 HUMAN
ID K2S3_HUMAN STANDARD; PRT; 304 AA.
AC Q14952; Q00644;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Killer cell immunoglobulin-like receptor 2DS3 precursor (MHC class I
DE NK cell receptor) (Natural killer associated transcript 7) (NKAT-7).
GN KIR2DS3 OR NKAT7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96270004; PubMed=8662091;
RA Doebering C., Samaridis J., Colonna M.;
RT "Alternatively spliced forms of human killer inhibitory receptors.";
RL Immunogenetics 44:227-230(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoid;
RA Biassoni R.;

```

Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: RECEPTOR ON NATURAL KILLER (NK) CELLS FOR HLA-C ALLELES.
-1- DOES NOT INHIBIT THE ACTIVITY OF NK CELLS.
-1- SUBCELLULAR LOCATION: Type I membrane protein.
-1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
-1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L76670; AAB36598.1; -
EMBL; X97231; CAA65870.1; -
HSP; P43626; INKR.
Genew; HGNC:6335; KIR2DS3.
MIM; 604954; -
InterPro; IPR003599; Ig.
InterPro; IPR003006; Ig_MHC.
InterPro; IPR003600; Ig_Like.
Pfam; PF00047; Ig; 2.
SMART; SM00409; Ig; 1.
SMART; SM00410; Ig_Like; 1.
Receptor; Immunoglobulin domain; Glycoprotein; Signal; Transmembrane;
Repeat; Multigene family.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 304 KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTOR
FT 2DS3.
FT DOMAIN 22 245 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 246 264 POTENTIAL.
FT DOMAIN 265 304 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 42 107 IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN 142 205 IG-LIKE C2-TYPE DOMAIN 2.
FT DISULFID 49 100 BY SIMILARITY.
FT DISULFID 149 198 BY SIMILARITY.
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 304 AA; 33717 MW; 80693F79844F9E7E CRC64;

Query Match 50.0%; Score 39; DB 1; Length 304;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 PKRMGGYRC 14
Db 189 PATQGGTYRC 198

RESULT 6
STCC EMENI
ID STCC EMENI STANDARD; PRT; 311 AA.
AC Q00668;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative sterigmatocystin biosynthesis peroxidase stcC precursor (EC 1.11.1.-).
GN STCC.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 26;
RX MEDLINE=96202293; PubMed=8643646;
RA Brown D.W., Yu J.-H., Kellar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;
RT "Twenty-five coregulated transcripts define a sterigmatocystin gene

cluster in Aspergillus nidulans.";
Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422 (1996).
-1- PATHWAY: Sterigmatocystin biosynthesis.
-1- SIMILARITY: BELONGS TO THE CHLOROPEROXIDASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U34740; AAC49193.1; -
InterPro; IPR000028; Chloroperoxidase.
Pfam; PF01328; Peroxidase_2; 1.
ProDom; PD040763; Chloroperoxidase; 1.
KW Oxidoreductase; Peroxidase; Iron; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 311 PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS
FT METAL 44 44 PEROXIDASE STCC.
FT METAL 44 44 IRON (PROTOHEME IX) (BY SIMILARITY).
SQ SEQUENCE 311 AA; 34816 MW; 18CB48B015CED735 CRC64;

Query Match 50.0%; Score 39; DB 1; Length 311;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 13
Db 215 ERLPVRKGGGAR 226

RESULT 7
IBMP CAMVP
ID IBMP CAMVP STANDARD; PRT; 520 AA.
AC P18617;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Inclusion body matrix protein (Viroplasmin).
GN VI.
OS Cauliflower mosaic virus (strain PV147) (CaMV).
OC Viruses; Retroid viruses; Caulimoviridae; Caulimovirus.
OX NCBI_TaxID=10647;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384848; PubMed=2402462;
RA Volovitch M., Modjtahedi N., Chouikh Y., Yot P.;
RT "DNA sequence of gene VI of cauliflower mosaic virus strain PV147.";
RL Nucleic Acids Res 18:5297-5297(1990).
CC -1- FUNCTION: ENHANCES THE TRANSLATION OF DOWNSTREAM ORF'S ON
CC POLYCISTRONIC MRNA'S DERIVED FROM CAULIFLOWER MOSAIC VIRUS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC INCLUSION BODIES.
CC -1- MISCELLANEOUS: THE INCLUSION BODIES ARE THE SITE OF VIRAL DNA
CC SYNTHESIS, VIRION ASSEMBLY AND ACCUMULATION IN THE INFECTED CELL.
CC -1- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES VIROPLASMIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X53860; CAA37853.1; -
PIR; S11217; S11217.
InterPro; IPR002609; Caulimo VI.
Pfam; PF01693; Caulimo_VI; 1.
KW Trans-acting factor; Translation regulation.
SQ SEQUENCE 520 AA; 57746 MW; 4CB77C155B5E1D65 CRC64;

```
Query Match 50.0%; Score 39; DB 1; Length 520;
Best Local Similarity 46.2%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYRC 14
DB 464 ERTVEKEGGSYKC 476

RESULT 8
IDH1_HUMAN STANDARD; PRT; 1319 AA.
AC Q10571;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable tumor suppressor protein MN1.
GN MN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95249266; PubMed=7731706;
RA Deprez R.H.L., Riegman P.H.J., Groen N.A., Warringa U.L.,
RA van Biesen N.A., Molijn A.C., Bootma D., de Jong P.J.,
RA Menon A.G., Kley N.A., Seizenger B.R., Zwarthoff E.C.,
RT "Cloning and characterization of MN1, a gene from chromosome 22q11,
RT which is disrupted by a balanced translocation in a meningioma.";
RL Oncogene 10:1521-1528(1995).
RN [2]
RP SEQUENCE OF 1304-1319 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97145634; PubMed=9026990;
RA Dmitrenko V.V., Garifulin O.M., Shostak E.A., Smikodub A.I.,
RA Kavsan V.M.;
RT "The characteristics of different types of mRNA expressed in the human
RT brain.";
RL Cyt. Genet. (Russ.) 30:41-47(1996).
CC -!- FUNCTION: MAY PLAY A ROLE IN TUMOR SUPPRESSION.
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED. HIGHEST LEVELS IN
CC SKELETAL MUSCLE.
CC -!- DISEASE: A FORM OF ACUTE MYELOID LEUKEMIA (AML) IS CHARACTERIZED
CC BY A CHROMOSOMAL TRANSLOCATION T(12;22)(P13;Q11) THAT INVOLVES
CC MN1 AND TEL.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-30 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X82209; CAA57693.1; ALT_INIT.
CC EMBL; Z70218; CAA94179.1; -.
CC Genew; HGNC:7180; MN1.
CC MIM; 156100; -.
CC Anti-oncogene; Chromosomal translocation; Alternative splicing.
FT DOMAIN 295 309 POLY-GLN.
FT DOMAIN 523 550 POLY-GLN.
SQ SEQUENCE 1319 AA; 135943 MW; 21197C9BBA272BE2 CRC64;

Query Match 50.0%; Score 39; DB 1; Length 1319;
Best Local Similarity 87.5%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 LPKRMGGS 11

Query Match 50.0%; Score 38; DB 1; Length 361;
Best Local Similarity 63.6%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSY 12
DB 20 QLLPKYGGRY 30

RESULT 10
GKIP2_HUMAN STANDARD; PRT; 553 AA.
ID GKIP2_HUMAN
AC Q14410;
```

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycerol kinase, testis specific 2 (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK).
 GN GKP2 OR GKTA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=95078834; PubMed=7987308;
 RA Sargent C.A., Young C., Marsh S., Ferguson-Smith M.A., Affara N.A.;
 RT "The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons";
 RL Hum. Mol. Genet. 3:1317-1324 (1994).
 CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.
 CC -1- PATHWAY: Glycerol utilization; rate-limiting step.
 CC -1- SUBCELLULAR LOCATION: BOUND TO THE MITOCHONDRIAL SURFACE OR CYTOPLASMIC. IN SPERM, THE MAJORITY OF THE ENZYME IS BOUND TO MITOCHONDRIA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCONOKINASE / GLYCEROKINASE / XYLULOXINASE FAMILY.
 CC -1- CAUTION: THIS COULD BE THE PRODUCT OF A PSEUDOGENE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).
 CC
 DR EMBL; X78712; CAA53365.1; -;
 DR HSPSP; P08859; IGLJ.
 DR Genew; HGNC:4291; GKP2.
 DR MIM; 600148; -;
 DR InterPro; IPR000577; FGGY_kin.
 DR Pfam; PF02782; FGGY_C1.
 DR TIGRFAMs; TIGR01311; glycerol_kin; 1.
 DR PROSITE; PS00445; FGGY_KINASES_2; 1.
 DR PROSITE; PS00933; FGGY_KINASES_1; 1.
 KW Glycerol metabolism; Transferase; Kinase; ATP-binding.
 FT NP BIND 167 179 ATP (PROBABLE).
 SQ SEQUENCE 553 AA; 60609 MW; 8CF53B1686BC4AD6 CRC64;
 Query Match 48.7%; Score 38; DB 1; Length 553;
 Best Local Similarity 53.8%; Pred. No. 33;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 ORLPKRMGGSYRC 14
 Db 346 BRLAKEVGTSTGC 358
 ||| :|||
 :||| :|||
 RESULT 11
 ID -SCAA CHICK STANDARD; PRT; 637 AA.
 AC Q92075; P70095; Q98941;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aniloride-sensitive sodium channel alpha-subunit (Epithelial Na+ channel alpha subunit) (Alpha ENaC) (Nonvoltage-gated sodium channel 1 alpha subunit) (SCNEA) (Alpha NaCh).
 GN SCN1A OR ENAC.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=Isa brown; TISSUE=Cochlea;
 RX MEDLINE=97157073; PubMed=9003454;
 RA Killick R., Richardson G.;
 RT "Isolation of chicken alpha ENaC splice variants from a cochlear cDNA library";
 RL Biochim. Biophys. Acta 1350:33-37 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Intestine;
 RA Goldstein O., Asher C., Garty H.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL INHIBITED BY THE DIURETIC AMILORIDE. MEDIATES THE ELECTRODIFFUSION OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY) THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
 CC ALSO PLAYS A ROLE IN TASTE PERCEPTION (BY SIMILARITY).
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA, ONE BETA AND ONE GAMMA SUBUNIT. A DELTA SUBUNIT CAN REPLACE THE ALPHA SUBUNIT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a short form, are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: THE LONG ISOFORM HAS BEEN FOUND IN COCHLEA, COLON, AND CARTILAGE. THE SHORT ISOFORM IS ONLY FOUND IN COCHLEA.
 CC -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNEL FAMILY.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-15 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).
 CC
 DR EMBL; U62902; AAB50550.1; -;
 DR EMBL; U62903; AAB50551.1; -;
 DR EMBL; U62904; AAB50552.1; -;
 DR EMBL; U58475; AAB04954.1; ALT_INIT.
 DR InterPro; IPR004724; Enac.
 DR InterPro; IPR001873; Na+channel_ASC.
 DR Pfam; PF00858; ASC; 1.
 DR PRINTS; PR01078; AMINACHANNEL.
 DR TIGRFAMs; TIGR00859; ENaC; 1.
 DR PROSITE; PS01206; ASC; 1.
 DR Ionic channel; Transmembrane; Ion transport; Glycoprotein;
 KW Alternative splicing.
 FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 81 97 POTENTIAL.
 FT DOMAIN 98 554 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 555 571 POTENTIAL.
 FT DOMAIN 572 637 POTENTIAL.
 FT CARBOHYD 157 157 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 298 298 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPPLIC 402 434 CIRCFCQANWKECSAVYFPLPDGAEVCDYT -> TCDH QFRASFFASFPGLQSPNSRACPRAMLR (IN SHORT ISOFORM).
 FT VARSPLIC 435 637 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 49 50 DV -> EL (IN REF. 2).
 FT CONFLICT 578 578 O -> R (IN REF. 2).
 SQ SEQUENCE 637 AA; 73236 MW; 5D15E616373971DA CRC64;
 Query Match 48.7%; Score 38; DB 1; Length 637;
 Best Local Similarity 50.0%; Pred. No. 39;

Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSY 12
 : : : : :
 Db 366 RKEMTERLGGSY 377

RESULT 12
 PRO-DROME
 ID PRO-DROME STANDARD; PRT; 1403 AA.
 AC Q29617; Q9U6A2; Q9VG98;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein prospero.
 GN PROS OR CG17228.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92069760; PubMed=1720353;
 RA Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
 RT "Prospero is expressed in neuronal precursors and encodes a nuclear
 RT protein that is involved in the control of axonal outgrowth in
 RT Drosophila.";
 RL Cell 67:941-953(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92171948; PubMed=1540176;
 RA Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
 RT "Cloning of the Drosophila prospero gene and its expression in
 RT ganglion mother cells.";
 RL Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93083413; PubMed=1842358;
 RA Chu-Lagraft O., Wright D.M., McNeil L.K., Doe C.Q.;
 RT "The prospero gene encodes a divergent homeodomain protein that
 RT controls neuronal identity in Drosophila.";
 RL Development Suppl. 2:79-85(1991).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).
 RX MEDLINE=20503846; PubMed=11051550;
 RA Xu C., Kaufmann R.C., Zhang J., Kladny S., Carthew R.W.;
 RT "Overlapping activators and repressors delimit transcriptional
 RT response to receptor tyrosine kinase signals in the Drosophila eye.";
 RL Cell 103:87-97(2000).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burkis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
 RA Chertis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [6]
 RP SIMILARITY TO C.ELEGANS CEH-26.
 RX MEDLINE=94212446; PubMed=7909177;
 RA Buerklin T.R.;
 RT "A Caenorhabditis elegans prospero homologue defines a novel domain.";
 RL Trends Biochem. Sci. 19:70-71(1994).
 CC -!- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR
 CC GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS
 CC CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED
 CC FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR
 CC PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT
 CC TRANSCRIPTION BY BINDING TO DNA.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: PROS-L (SHOWN HERE) AND PROS-
 CC S; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-
 CC TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.
 CC -!- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE
 CC DEVELOPING CNS, LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING
 CC FORMATION.
 CC -!- SIMILARITY: BELONGS TO THE PROSERO HOMEBOX FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81389; AAA28841.1; -;
 CC EMBL; D10609; BAA01464.1; -;
 CC EMBL; Z11743; CAA7802.1; -;
 CC EMBL; AF190403; AAF05703.1; -;
 CC EMBL; AE003691; AAF54628.1; ALT_INIT.
 CC PIR; A41089; A41089.
 CC PIR; JQ1397; JQ1397.
 CC FlyBase; FBgn004595; pros.
 CC Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
 CC Developmental protein; Alternative splicing.
 CC DOMAIN 4 12 POLY-ALA.
 CC DOMAIN 28 31 POLY-SER.
 CC DOMAIN 32 35 POLY-ASN.
 CC DOMAIN 188 191 POLY-ALA.
 CC DOMAIN 216 264 GLN-RICH.
 CC DOMAIN 270 286 ASN-RICH.
 CC DOMAIN 318 354 SER-RICH.
 CC DOMAIN 431 437 POLY-RICH.
 CC DOMAIN 505 508 POLY-ASP.
 CC DOMAIN 700 1048 POLY-ALA.
 CC DOMAIN 934 937 GLN-RICH.
 CC DOMAIN 991 998 POLY-ALA.
 CC DOMAIN 1074 1082 NUCLEAR LOCALIZATION SIGNAL.
 CC POLY-PRO.

```

FT DOMAIN 1127 1137 HIS-RICH.
FT DNA_BIND 1241 1303 HOMEOBOX (ATYPICAL).
FT DOMAIN 1304 1403 PROSPERO-LIKE.
FT VARSPLIC 1216 1244 MISSING (IN ISOFORM PROS-S).
FT CONFLICT 76 98 AKMLNELFGKQKQADATSGLP -> GQDAERAVWPPDEA
FT CONFLICT 120 144 GGRNEMPA (IN REF. 1 AND 4).
FT CONFLICT 418 418 H -> Q (IN REF. 2).
FT CONFLICT 677 677 A -> C (IN REF. 1).
FT CONFLICT 802 802 A -> S (IN REF. 1).
FT CONFLICT 958 958 T -> S (IN REF. 1).
FT CONFLICT 1048 1048 Q -> QQQQ (IN REF. 1).
SQ SEQUENCE 1403 AA; 153569 MW; 95FB9973E24E238E CRC64;

Query Match 48.7%; Score 38; DB 1; Length 1403;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGS 11
||| |||||:
Db 794 RLPTRMGGA 802

RESULT 13
PRO_DROVI STANDARD; PRT; 1556 AA.
AC Q9U6A1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein prospero.
GN PROS.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=20503846; PubMed=11051550;
RA Xu C., Kauffmann R.C., Zhang J., Kladny S., Carthew R.W.;
RX "Overlapping activators and repressors delimit transcriptional
RT response to receptor tyrosine kinase signals in the Drosophila eye.";
RL Cell 103:87-97(2000).
CC -1- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR
CC GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS
CC CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED
CC FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR
CC PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT
CC FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE
CC TRANSCRIPTION BY BINDING TO DNA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -1- SIMILARITY: BELONGS TO THE PROSPERO HOMEOBOX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF190405; AAF06660.1; -.
CC FlyBase; FBgn028753; Dvir1pros.
CC Nuclear protein; Transcription regulation; DNA-binding;
CC Homeobox; Developmental protein.
CC
CC DOMAIN 4 12 POLY-ALA.
CC FT DOMAIN 19 36 ASN-RICH.
CC FT DOMAIN 150 172 ASN-RICH.
CC FT DOMAIN 206 209 POLY-ALA.
CC FT DOMAIN 237 292 GLN-RICH.

```

```

FT DOMAIN 305 309 POLY-ASN.
FT SER-RICH. 349 381
FT ASP-RICH. 443 473
FT POLY-ALA. 544 547
FT GLN-RICH. 574 1080
FT ASN-RICH. 888 906
FT DOMAIN 1027 1030 POLY-ALA.
FT DOMAIN 1045 1054 POLY-GLN.
FT DOMAIN 1057 1062 POLY-GLN.
FT THR-RICH. 1132 1189
FT POLY-ALA. 1140 1145
FT DOMAIN 1154 1163 POLY-GLN.
FT DOMAIN 1183 1189 POLY-THR.
FT DOMAIN 1090 1097 NUCLEAR LOCALIZATION SIGNAL (BY
SIMILARITY).
FT POLY-GLN. 1330 1337
FT HOMEOBOX (ATYPICAL) (BY SIMILARITY).
FT DOMAIN 1457 1556 PROSPERO-LIKE (BY SIMILARITY).
SQ SEQUENCE 1556 AA; 171029 MW; 6FEACFEA2D73E644 CRC64;

Query Match 48.7%; Score 38; DB 1; Length 1556;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGS 11
||| |||||:
Db 853 RLPTRMGGA 861

RESULT 14
RS14_AERPE STANDARD; PRT; 54 AA.
ID RS14_AERPE
AC P58731;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S14P.
GN RPS14P OR APE0353.1.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococccaceae; Aeropyrum.
OX NCBI_TaxID=56636;
[1]
SEQUENCE FROM N.A.
RP STRAIN=K1;
RC MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -1- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF000059; -. NOT ANNOTATED CDS.
CC InterPro; IPR001209; Ribosomal_S14.
CC Pfam; PF00253; Ribosomal_S14; 1.
CC PROSITE; PS00527; RIBOSOMAL_S14; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 54 AA; 6326 MW; EB7FF17E9205C0CD CRC64;

Query Match 48.1%; Score 37.5; DB 1; Length 54;

```


Best Local Similarity 72.7%; Pred. No. 3.8;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 5 PKRMG-GSYRC 14
| | | | | : | |
Db 9 PKRMGGAQRC 19

RESULT 15

IE68_HSV2
ID IE68_HSV2 STANDARD; PRT; 197 AA.
AC P14379;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Immediate-early protein IE4 (IE68) (Fragment).
GN US1.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10310;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84137573; PubMed=6321634;
RA Whitton J.L., Clements J.B.;
RT "The junctions between the repetitive and the short unique sequences
RT of the herpes simplex virus genome are determined by the polypeptide-
RT coding regions of two spliced immediate-early mRNAs.";
RL J. Gen. Virol. 65:451-466(1984)
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC HSV-2 IE-68 (US1), EHV-1 65, EHV-4 (ORF4), PRV RSP40, AND VZV 63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29384; AAA45848.1; -.
KW Early protein.
FT NON_TER 197
SQ SEQUENCE 197 AA; 21510 MW; 314C23F55C795CBB CRC64;

Query Match 47.4%; Score 37; DB 1; Length 197;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSY 12
| : | : | |
Db 68 RRRPPRELGGRY 79

Search completed: March 4, 2003, 16:53:42
Job time : 5.27119 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:51:50 ; Search time 35.8305 Seconds
(without alignments)
80.508 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RRLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	42	53.8	107	2 O68807	O68807 synchococc
2	42	53.8	339	5 Q9VUB3	Q9vub3 drosophila
3	41	52.6	83	16 Q8YZU7	Q8yzu7 anabaena sp
4	41	52.6	564	16 Q9RUK9	Q9ruk9 deinococcus
5	40.5	51.9	235	11 Q9J127	Q9j127 rattus norv
6	40.5	51.9	235	11 Q9J126	Q9j126 rattus norv
7	40	51.3	151	16 Q9K4D3	Q9k4d3 streptomyce
8	40	51.3	220	16 Q8X6I5	Q8x6i5 escherichia
9	40	51.3	314	16 Q9RWV7	Q9rwv7 deinococcus
10	40	51.3	413	10 Q94EC2	Q94ec2 oryza sativ
11	40	51.3	472	10 Q8RZS4	Q8rzs4 oryza sativ
12	40	51.3	538	2 Q9Z4V9	Q9z4v9 streptomyce
13	40	51.3	559	4 Q969Y0	Q969y0 homo sapien
14	40	51.3	1556	10 Q9NB97	Q9nb97 nicotiana t
15	40	51.3	1559	10 Q49889	Q49889 lycopersico
16	40	51.3	1955	10 Q8W078	Q8w078 oryza sativ

17	39	50.0	61	12	Q83177	Q83177 cauliflower
18	39	50.0	212	12	Q9YP05	Q9yp05 choriatoneu
19	39	50.0	298	17	O50108	O50108 pyrococcus
20	39	50.0	332	16	Q9PI09	Q9pi09 campylobact
21	39	50.0	372	17	Q97U79	Q97u79 sulfolobus
22	39	50.0	395	16	Q8U8A2	Q8u8a2 agrobacteri
23	39	50.0	472	2	Q9L3E9	Q9l3e9 amycolatops
24	39	50.0	520	12	Q83170	Q83170 cauliflower
25	39	50.0	520	12	Q66163	Q66163 cauliflower
26	39	50.0	542	11	O70334	O70334 mus musculu
27	39	50.0	575	4	Q9NXQ2	Q9nxq2 homo sapien
28	39	50.0	992	5	Q9VM91	Q9vm91 drosophila
29	39	50.0	1260	4	Q9UGF9	Q9ugf9 homo sapien
30	39	50.0	1533	4	O75560	O75560 homo sapien
31	39	50.0	1677	11	O70373	O70373 mus musculu
32	39	50.0	1862	11	Q925B2	Q925b2 mus musculu
33	39	50.0	1863	11	Q9JLQ1	Q9jlq1 mus musculu
34	39	50.0	1863	11	Q923J1	Q923j1 mus musculu
35	39	50.0	1864	4	Q9BXB2	Q9bxb2 homo sapien
36	39	50.0	1865	4	Q96QT4	Q96qt4 homo sapien
37	39	50.0	2488	5	O61528	O61528 caenorhabdi
38	38.5	49.4	130	11	Q9D5C1	Q9d5c1 mus musculu
39	38	48.7	210	2	Q53166	Q53166 rhizobium s
40	38	48.7	299	5	Q8T1F2	Q8t1f2 dictyosteli
41	38	48.7	356	16	P96450	P96450 rhizobium m
42	38	48.7	359	10	Q94DU7	Q94du7 oryza sativ
43	38	48.7	372	17	Q976X2	Q976x2 sulfolobus
44	38	48.7	385	2	Q9ZN85	Q9zn85 corynebacte
45	38	48.7	391	10	Q9LVZ9	Q9lvz9 arabidopsis

ALIGNMENTS

RESULT 1

O68807 PRELIMINARY; PRT; 107 AA.

ID O68807

AC O68807; (T=EMBLrel. 07, Created)

DT 01-AUG-1998 (T=EMBLrel. 07, Last sequence update)

DT 01-AUG-1998 (T=EMBLrel. 07, Last annotation update)

DE Hypothetical 12.2 kDa protein (fragment).

OS Synchococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).

OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.

OX NCBI_TaxID=32049;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PCC7002;

RA Droog F.N.J., Tallier B.J., Stevens S.E. Jr.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF054515; AAC08034.1; -

KW Hypothetical protein

FT NON_TER 107

SQ SEQUENCE 107 AA; 12163 MW; A5A44929EB610569 CRC64;

Query Match 53.8%; Score 42; DB 2; Length 107;
Best Local Similarity 66.7%; Pred. No. 4.1;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 13

||| ||| |||
||| ||| |||

RESULT 2

Q9VUB3 PRELIMINARY; PRT; 339 AA.

ID Q9VUB3

AC Q9VUB3; (T=EMBLrel. 13, Created)

DT 01-MAY-2000 (T=EMBLrel. 13, Last sequence update)

DT 01-JUN-2002 (T=EMBLrel. 21, Last annotation update)

DE CG17359 protein (RE58063p).

GN CG17359.

OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh A.A., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celinker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003536; AAF49775.1; -;
 DR EMBL; AY071545; AAL49167.1; -;
 DR HSP; PB0045; 12NF.
 DR FlyBase; Fg00036396; CG17359.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 4.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 KW DNA-binding; Metal-binding; Zinc-binding; Zinc-finger.
 SQ SEQUENCE 339 AA; 39250 MW; 98653D457205C57A CRC64;

Query Match 53.8%; Score 42; DB 5; Length 339;
 Best Local Similarity 70.0%; Pred. No. 14;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PKRMGGSYRC 14
 |||:||||
 Db 208 PKRVPGPYRC 217

RESULT 3

Q8YZU7 PRELIMINARY; PRT; 83 AA.
 ID Q8YZU7
 AC Q8YZU7
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein Asr0358.
 GN ASR0358.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Matanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimoto S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120.";
 RL DNA Res. 8:2105-213(2001).
 DR EMBL; AP003582; BAB72316.1; -;
 DR InterPro; IPR003594; ATPbind_ATPase.
 DR Pfam; PF02518; HATPase_c1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 83 AA; 9321 MW; B315C9D52BDA04EC CRC64;

Query Match 52.6%; Score 41; DB 16; Length 83;
 Best Local Similarity 38.5%; Pred. No. 4.8;
 Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYRC 14

: : |||:||||
 Db 46 RNTAKLGGNFR 58

RESULT 4

ID Q9RUK9 PRELIMINARY; PRT; 564 AA.
 AC Q9RUK9
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Glycosyl hydrolase, family 13.
 GN Drl375.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcaceae.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Mathavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium *Deinococcus*
radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001983; AAF10944.1; -;
 DR HSSP; P21332; IUOK.

```
DR TTGR; DRI375; -.
DR InterPro; IPR000461; Alpha_amylase.
DR Pfam; PF00128; alpha_amylase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 564 AA; 63667 MW; B8F50B9B0DFC8D51 CRC64;

Query Match 52.6%; Score 41; DB 16; Length 564;
Best Local Similarity 61.5%; Pred. No. 38;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 RRLPKRMGGSYR 13
DB 487 RQEPALVGGSYR 499

RESULT 5
Q9J127 ID Q9J127 PRELIMINARY; PRT; 235 AA.
AC Q9J127;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TNF-alpha propeptide 5 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=DARK AGOUTI;
RA Seidel M.F., Junier M.-P., Vetter H.;
RT "TNF-alpha polymorphism in rats with collagen-induced arthritis.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF269159; AAF82567.1; -.
DR HSSP; P06804; 2TNF.
DR InterPro; IPR003636; TNF_abc.
DR InterPro; IPR000478; TNF_family.
DR Pfam; PF00229; TNF; 1.
DR PRINTS; PR01234; TNECROSISFCT.
DR ProDom; PD002012; TNF_abc; 1.
DR SMART; SM00207; TNF; 1.
DR PROSITE; PS00251; TNF_1; 1.
DR PROSITE; PS50049; TNF_2; 1.
FT NON TER 235
SQ SEQUENCE 235 AA; 25806 MW; 12A8EC6D0491428D CRC64;

Query Match 51.9%; Score 40.5; DB 11; Length 235;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14
DB 15 EALPKRMGGGLQNSRR 30

RESULT 6
Q9J126 ID Q9J126 PRELIMINARY; PRT; 235 AA.
AC Q9J126;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TNF-alpha propeptide 3 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=DARK AGOUTI;
RA Seidel M.F., Junier M.-P., Vetter H.;
RT "TNF-alpha polymorphism in rats with collagen-induced arthritis.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF269159; AAF82567.1; -.
DR HSSP; P06804; 2TNF.
DR InterPro; IPR003636; TNF_abc.
DR InterPro; IPR000478; TNF_family.
DR Pfam; PF00229; TNF; 1.
DR PRINTS; PR01234; TNECROSISFCT.
DR ProDom; PD002012; TNF_abc; 1.
DR SMART; SM00207; TNF; 1.
DR PROSITE; PS00251; TNF_1; 1.
DR PROSITE; PS50049; TNF_2; 1.
FT NON TER 235
SQ SEQUENCE 235 AA; 25806 MW; 12A8EC6D0491428D CRC64;

Query Match 51.9%; Score 40.5; DB 11; Length 235;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14
DB 15 EALPKRMGGGLQNSRR 30

RESULT 7
Q9K4D3 ID Q9K4D3 PRELIMINARY; PRT; 151 AA.
AC Q9K4D3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted/membrane protein.
GN SC05375 OR 2SC65.19.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL359152; CAB94546.1; -.
SQ SEQUENCE 151 AA; 16639 MW; 590A5ACC6104812B CRC64;

Query Match 51.3%; Score 40; DB 16; Length 151;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLPKRMGGSYRC 14
DB 27 RRLIQRSRGTFDC 40

RESULT 8
Q8X615 ID Q8X615 PRELIMINARY; PRT; 230 AA.
AC Q8X615;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Orf, hypothetical protein.
GN YQEH OR 24166 OR ECS3703.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
DR EMBL; AF269160; AAF82568.1; -.
DR HSSP; P06804; 2TNF.
DR InterPro; IPR003636; TNF_abc.
DR InterPro; IPR000478; TNF_family.
DR Pfam; PF00229; TNF; 1.
DR PRINTS; PR01234; TNECROSISFCT.
DR ProDom; PD002012; TNF_abc; 1.
DR SMART; SM00207; TNF; 1.
DR PROSITE; PS00251; TNF_1; 1.
DR PROSITE; PS50049; TNF_2; 1.
FT NON TER 235
SQ SEQUENCE 235 AA; 25789 MW; C801B92D049C2F2E CRC64;

Query Match 51.9%; Score 40.5; DB 11; Length 235;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY 2 QRLPKRMGG---SYRC 14
DB 15 EALPKRMGGGLQNSRR 30

RESULT 7
Q9K4D3 ID Q9K4D3 PRELIMINARY; PRT; 151 AA.
AC Q9K4D3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted/membrane protein.
GN SC05375 OR 2SC65.19.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL359152; CAB94546.1; -.
SQ SEQUENCE 151 AA; 16639 MW; 590A5ACC6104812B CRC64;

Query Match 51.3%; Score 40; DB 16; Length 151;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLPKRMGGSYRC 14
DB 27 RRLIQRSRGTFDC 40

RESULT 8
Q8X615 ID Q8X615 PRELIMINARY; PRT; 230 AA.
AC Q8X615;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Orf, hypothetical protein.
GN YQEH OR 24166 OR ECS3703.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EBL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouais K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005513; AAG57958.1; -.
DR ENBL; AP002563; BAB37126.1; -.
DR InterPro; IPR000792; HTH_LuxR.
KW Complete proteome.
SQ SEQUENCE 230 AA; 26730 MW; C61F978B70FF13D4 CRC64;

Query Match 51.3%; Score 40; DB 16; Length 230;
Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGG 10
Db 207 KQRIKRMGG 216

RESULT 9
Q9RWM7 PRELIMINARY; PRT; 314 AA.
AC Q9RWM7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Pilin, type IV, putative.
GN DR0548.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Matthevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -1- SUBUNIT: THE PILI ARE POLAR FLEXIBLE FILAMENTS OF ABOUT 5.4
CC NANOMETERS DIAMETER AND 2.5 MICROMETERS AVERAGE LENGTH; THEY
CC CONSIST OF ONLY A SINGLE POLYPEPTIDE CHAIN ARRANGED IN A HELICAL
CC CONFIGURATION OF FIVE SUBUNITS PER TURN IN THE ASSEMBLED PILUS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE N-ME-PHE PILI FAMILY.
```

```
DR EMBL; AB001913; AAF10127.1; -.
DR TIGR; DR0548; -.
DR InterPro; IPR001082; Pilin.
DR InterPro; IPR001120; Prok_N_methyltn.
DR Pfam; PF00114; pilin; 1.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Fimbria; Methylation; Complete proteome.
SQ SEQUENCE 314 AA; 31867 MW; 96619CSB005A3B86 CRC64;

Query Match 51.3%; Score 40; DB 16; Length 314;
Best Local Similarity 72.7%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGG 11
Db 26 RQRGPRQGG 36

RESULT 10
Q94EC2 PRELIMINARY; PRT; 413 AA.
AC Q94EC2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE P0002B05.17 protein.
GN P0002B05.17.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0002B05.17."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003141; BAB63497.1; -.
DR InterPro; IPR001841; Znf_ring.
SQ SEQUENCE 413 AA; 43766 MW; 741AEFDC1C77EB13 CRC64;

Query Match 51.3%; Score 40; DB 10; Length 413;
Best Local Similarity 57.1%; Pred. No. 41;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14
Db 42 RQRVKQRGGGGRG 55

RESULT 11
Q8RZS4 PRELIMINARY; PRT; 472 AA.
AC Q8RZS4
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative OSGA2ox1.
GN B1140D12.2.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
RT clone: B1140D12."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL: AP003537; BAB86157.1; -.
SQ SEQUENCE 472 AA; 49710 MW; 1B50411111B123139 CRC64;

Query Match          51.3%; Score 40; DB 10; Length 472;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 RRLPKRMGGSYRC 14
   :|||:|||:|||
DB 241 QRLRRWGLWRC 254

RESULT 12
Q924V9 PRELIMINARY; PRT; 538 AA.
AC Q924V9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE NAD(+) synthase (Glutamine-hydrolysing) (EC 6.3.5.1) (Fragment).
GN NADE.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=99445172; PubMed=10517584;
RA Fink D., Falke D., Wohlleben W., Engels A.;
RT "Nitrogen metabolism in Streptomyces coelicolor A3(2): modification of
RL glutamine synthetase I by an adenyltransferase.";
RL Microbiology 145:2313-2322(1999).
DR EMBL: Y17736; CAB38325.1; -.
DR InterPro: IPR003694; NAD_synthase.
DR InterPro: IPR003010; Ntase/CNhydase.
DR Pfam: PF02540; NAD_synthase; 1.
DR TIGRFAMs: TIGR00552; nade; 1.
KW Ligase.
FT NON_TER
FT NON_TER
SQ SEQUENCE 538 AA; 57986 MW; B17D1AD0EA833AA CRC64;

Query Match          51.3%; Score 40; DB 2; Length 538;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PKRMGGSYRC 14
   |||:|||
DB 233 PGRLLGSSRC 242

RESULT 13
Q969Y0 PRELIMINARY; PRT; 559 AA.
AC Q969Y0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA FLJ30102 fis, clone BNGH41000137, weakly similar to brush border
DE 61.9 kDa protein precursor (Unknown) (protein for MGC:15606).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsubo T., Sato H., Ota T., Wakamatsu A.,
RA Lehl S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
```

```
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK054664; BAB70787.1; -.
DR EMBL: BC009431; AAH09431.1; -.
SQ SEQUENCE 559 AA; 63799 MW; 56264FEEBFCF358 CRC64;

Query Match          51.3%; Score 40; DB 4; Length 559;
Best Local Similarity 63.6%; Pred. No. 57;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSY 12
   |||:|||
DB 133 QRAPKYGDDY 143

RESULT 14
Q9WB97 PRELIMINARY; PRT; 1556 AA.
AC Q9WB97;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Modification methylase (EC 2.1.1.73) (Cytosine-specific
DE methyltransferase).
DE NMET1.
GN Nicotiana tabacum (Common tobacco).
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakano Y., Steward N., Sekine M., Kusano T., Sano H.;
RT "A tobacco NtMET1 cDNA, encoding a DNA methyltransferase: Molecular
RT characterization and abnormal phenotypes of transgenic tobacco
RT plants.";
RL Plant Cell Physiol. 0:0-0(2000).
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
DR EMBL: AB030726; BAA92852.1; -.
DR HSSP: OI4717; IG55.
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR Pfam: PF01426; BAH; 2.
DR Pfam: PF00145; DNA_methylase; 1.
DR PRINTS: PR00105; C5METTRFRASE.
DR SMART: SM00439; BAH; 2.
DR TIGRFAMs: TIGR00675; dcm; 1.
DR PROSITE: PS00094; C5_MTASE_1; 1.
DR PROSITE: PS00095; C5_MTASE_2; 1.
KW Methyltransferase; Transferase.
SQ SEQUENCE 1556 AA; 175545 MW; E4E3E296E842877A CRC64;

Query Match          51.3%; Score 40; DB 10; Length 1556;
Best Local Similarity 58.3%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 13
   |||:|||
DB 1420 QRIPKPGADWR 1431

RESULT 15
O49889 PRELIMINARY; PRT; 1559 AA.
ID O49889
AC O49889;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
```

DT 01-JUN-1998 (TREMBREL. 06, Last sequence update)
DT 01-JUN-2002 (TREMBREL. 21, Last annotation update)
DE Modification methylase (EC 2.1.1.73) (Cytosine-specific
DE methyltransferase).
GN LESMET.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MERAUVIGLIA; TISSUE=APICAL MERISTEMS;
RA Bernacchia G., Para A., Pedrali-Noy G., Cella R.;
RT "Isolation of a cDNA coding for DNA (Cytosine-5)-methyltransferase
RT (Accession No. AJ002140) from Lycopersicon esculentum (PGR98-008).";
RL Plant Physiol. 116:446-446(1998).
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC -!- ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
DR EMBL; AJ002140; CAA05207.1; -.
DR HSSP; O14717; IC55.
DR REBASE; 3251; M.LesIP.
DR InterPro; IPR001025; BAH.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF01426; BAH; 2.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PR00105; C5METTRFRASE.
DR SMART; SM00439; BAH; 2.
DR TIGRFAMS; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00095; C5_MTASE_2; 1.
KW Methyltransferase; Transferase.
SQ SEQUENCE 1559 AA; 174915 MW; F19ABED29B832C14 CRC64;

Query Match 51.3%; Score 40; DB 10; Length 1559;
Best Local Similarity 58.3%; Pred.No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ORLPKRMGGSYR 13
| | | | | : |
Db 1423 QRIPKPGADWR 1434

Search completed: March 4, 2003, 16:56:28
Job time : 45.8305 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:52:45 ; Search time 5.69491 Seconds
(without alignments)
72.331 Million cell updates/sec

Title: US-09-745-008-14
Perfect score: 78
Sequence: 1 RORLPKRMGGSYRC 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	50.0	542	1	US-08-412-431-3
2	39	50.0	542	1	US-08-623-679-3
3	39	50.0	542	3	US-08-933-774-3
4	39	50.0	542	4	US-09-181-030-3
5	39	50.0	542	4	US-09-534-242-3
6	39	50.0	542	4	US-09-454-854-3
7	39	50.0	542	4	US-09-164-671-3
8	39	50.0	1497	1	US-08-623-679-7
9	39	50.0	1497	3	US-08-933-774-7
10	39	50.0	1497	4	US-09-181-030-7
11	39	50.0	1497	4	US-09-534-242-7
12	39	50.0	1497	4	US-09-454-854-7
13	39	50.0	1497	4	US-09-164-671-7
14	39	50.0	1533	1	US-08-623-679-9
15	39	50.0	1533	3	US-08-933-774-9
16	39	50.0	1533	4	US-09-181-030-9
17	39	50.0	1533	4	US-09-534-242-9
18	39	50.0	1533	4	US-09-454-854-9
19	39	50.0	1533	4	US-09-164-671-9
20	37	47.4	146	4	US-09-522-217-56
21	37	47.4	192	4	US-09-134-001C-3263
22	37	47.4	341	1	US-08-248-628A-2
23	37	47.4	486	2	US-08-432-016-6
24	37	47.4	486	2	US-08-684-594-6
25	37	47.4	510	4	US-09-522-217-89
26	36.5	46.8	235	4	US-08-883-086-7
27	36	46.2	92	4	US-09-366-887A-14

28	36	46.2	357	4	US-09-134-001C-3379	Sequence 3379, Ap
29	35.5	45.5	560	2	US-08-808-931-18	Sequence 18, Appl
30	35.5	45.5	560	3	US-08-808-323-18	Sequence 18, Appl
31	35.5	45.5	560	3	US-09-050-603A-18	Sequence 18, Appl
32	35.5	45.5	560	3	US-09-102-420B-18	Sequence 18, Appl
33	35.5	45.5	560	4	US-09-497-698-18	Sequence 18, Appl
34	35	44.9	175	2	US-08-997-080-53	Sequence 53, Appl
35	35	44.9	175	2	US-08-997-362-53	Sequence 53, Appl
36	35	44.9	175	3	US-08-873-970-53	Sequence 53, Appl
37	35	44.9	175	4	US-09-095-855-53	Sequence 53, Appl
38	35	44.9	175	4	US-08-705-347A-53	Sequence 53, Appl
39	35	44.9	175	4	US-09-324-542-53	Sequence 53, Appl
40	35	44.9	175	4	US-09-205-426-53	Sequence 53, Appl
41	35	44.9	175	4	US-09-200-643-53	Sequence 53, Appl
42	35	44.9	341	4	US-08-853-948B-4	Sequence 4, Appl
43	35	44.9	507	1	US-08-484-493-12	Sequence 12, Appl
44	35	44.9	507	1	US-08-484-494-12	Sequence 12, Appl
45	35	44.9	507	2	US-08-345-212-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-412-431-3
; Sequence 3, Application US/08412431
; Patent No. 5633161
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/412,431
; FILING DATE: 29-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-412-431-3

Query Match 50.0%; Score 39; DB 1; Length 542;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
|:|:|:|:
Db 470 MGGAYRC 476

```
RESULT 2
US-08-623-679-3
; Sequence 3, Application US/08623679
; Patent No. 5674739
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR
; TITLE OF INVENTION: PROGRESSION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/623,679
; FILING DATE: 29-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/412,431
; FILING DATE: 29-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-623-679-3

Query Match 50.0%; Score 39; DB 1; Length 542;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
Db 470 MGGAYRC 476

RESULT 3
US-08-933-774-3
; Sequence 3, Application US/08933774A
; Patent No. 6025137
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004003
; CURRENT APPLICATION NUMBER: US/08/933,774A
; CURRENT FILING DATE: 1997-09-19
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; EARLIER APPLICATION NUMBER: US 08/412,431
; EARLIER FILING DATE: 1995-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mus musculus
;
US-08-933-774-3

Query Match 50.0%; Score 39; DB 4; Length 542;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
Db 470 MGGAYRC 476

RESULT 4
US-09-181-030-3
; Sequence 3, Application US/09181030
; Patent No. 6251597
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004005
; CURRENT APPLICATION NUMBER: US/09/181,030
; CURRENT FILING DATE: 1998-10-27
; EARLIER APPLICATION NUMBER: US 08/862,442
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mus musculus
;
US-09-181-030-3

Query Match 50.0%; Score 39; DB 4; Length 542;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
Db 470 MGGAYRC 476

RESULT 5
US-09-534-242-3
; Sequence 3, Application US/09534242
; Patent No. 6312909
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004004
; CURRENT APPLICATION NUMBER: US/09/534,242
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: US 09/164,671
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: US 08/862,442
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mus musculus
;
US-09-534-242-3

Query Match 50.0%; Score 39; DB 4; Length 542;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
Db 470 MGGAYRC 476
```

Qy 8 MGSYRC 14
 Db 470 MGGAYRC 476

RESULT 6

US-09-454-854-3
 ; Sequence 3, Application US/09454854

Patent No. 6316204

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION

AND TREATMENT OF TUMOR PROGRESSION

FILE REFERENCE: 07334/004005

CURRENT APPLICATION NUMBER: US/09/454,854

CURRENT FILING DATE: 1999-12-07

PRIOR APPLICATION NUMBER: 09/181,030

PRIOR FILING DATE: 1998-10-27

PRIOR APPLICATION NUMBER: US 08/623,679

PRIOR FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 3

LENGTH: 542

TYPE: PRT

ORGANISM: Mus musculus

US-09-454-854-3

Query Match 50.0%; Score 39; DB 4; Length 542;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
 Db 470 MGGAYRC 476

RESULT 7

US-09-164-671-3

; Sequence 3, Application US/09164671A

Patent No. 6372896

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION

AND TREATMENT OF TUMOR PROGRESSION

FILE REFERENCE: 07334/004004

CURRENT APPLICATION NUMBER: US/09/164,671A

CURRENT FILING DATE: 1998-10-01

EARLIER APPLICATION NUMBER: US 08/862,442

EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: US 08/623,679

EARLIER FILING DATE: 1996-03-29

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 3

LENGTH: 542

TYPE: PRT

ORGANISM: Mus musculus

US-09-164-671-3

Query Match 50.0%; Score 39; DB 4; Length 542;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
 Db 470 MGGAYRC 476

RESULT 8

US-08-623-679-7

; Sequence 7, Application US/08623679

Patent No. 5674739

GENERAL INFORMATION:
 APPLICANT: Shyjan, Andrew W.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR
 PROGRESSION
 TITLE OF INVENTION: PROGRESSION
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/623,679

FILING DATE: 29-MAR-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/412,431

FILING DATE: 29-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fasse, J. Peter

REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 07334/004001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1497 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-623-679-7

Query Match 50.0%; Score 39; DB 1; Length 1497;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGSYRC 14
 Db 471 MGGAYRC 477

RESULT 9

US-08-933-774-7

; Sequence 7, Application US/08933774A

Patent No. 6025137

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION

AND TREATMENT OF TUMOR PROGRESSION

FILE REFERENCE: 07334/004003

CURRENT APPLICATION NUMBER: US/08/933,774A

CURRENT FILING DATE: 1997-09-19

EARLIER APPLICATION NUMBER: US 08/623,679

EARLIER FILING DATE: 1996-03-29

EARLIER APPLICATION NUMBER: US 08/412,431

EARLIER FILING DATE: 1995-03-29

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSEQ for Windows Version 3.0

SEQ ID NO 7

LENGTH: 1497

TYPE: PRT

ORGANISM: Homo sapiens

US-08-933-774-7

Query Match 50.0%; Score 39; DB 3; Length 1497;

```
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
Db 471 MGGAYRC 477

RESULT 10
US-09-181-030-7
; Sequence 7, Application US/09181030
; Patent No. 6251597
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004005
; CURRENT APPLICATION NUMBER: US/09/181,030
; CURRENT FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: US 08/862,442
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-181-030-7

Query Match 50.0%; Score 39; DB 4; Length 1497;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
Db 471 MGGAYRC 477

RESULT 11
US-09-534-242-7
; Sequence 7, Application US/09534242
; Patent No. 6313909
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004004
; CURRENT APPLICATION NUMBER: US/09/534,242
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: US 09/164,671
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: US 08/862,442
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-534-242-7

Query Match 50.0%; Score 39; DB 4; Length 1497;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
Db 471 MGGAYRC 477

RESULT 12
US-09-454-854-7
; Sequence 7, Application US/09454854
; Patent No. 6316204
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004005
; CURRENT APPLICATION NUMBER: US/09/454,854
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/181,030
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: US 08/623,679
; PRIOR FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-454-854-7

Query Match 50.0%; Score 39; DB 4; Length 1497;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
Db 471 MGGAYRC 477

RESULT 13
US-09-164-671-7
; Sequence 7, Application US/09164671A
; Patent No. 6372896
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
; TITLE OF INVENTION: AND TREATMENT OF TUMOR PROGRESSION
; FILE REFERENCE: 07334/004004
; CURRENT APPLICATION NUMBER: US/09/164,671A
; CURRENT FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: US 08/862,442
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: US 08/623,679
; EARLIER FILING DATE: 1996-03-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-164-671-7

Query Match 50.0%; Score 39; DB 4; Length 1497;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYRC 14
Db 471 MGGAYRC 477

RESULT 14
US-08-623-679-9
; Sequence 9, Application US/08623679
; Patent No. 5674739
; GENERAL INFORMATION:
; APPLICANT: Shvjan, Andrew W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR
```

; TITLE OF INVENTION: PROGRESSION
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/623,679
 ; FILING DATE: 29-MAR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/412,431
 ; FILING DATE: 29-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fasse, J. Peter
 ; REGISTRATION NUMBER: 32,983
 ; REFERENCE/DOCKET NUMBER: 07334/004001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1533 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-623-673-9

Query Match 50.0%; Score 39; DB 1; Length 1533;
 Best Local Similarity 85.7%; Pred. NO. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 MGGSYRC 14
 DB 471 MGGAYRC 477

RESULT 15
 US-08-933-774-9
 ; Sequence 9, Application US/08933774A
 ; Patent No. 6025137
 ; GENERAL INFORMATION:
 ; APPLICANT: Shv'jan, Andrew W.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION
 ; FILE REFERENCE: 07334/004003
 ; CURRENT APPLICATION NUMBER: US/08/933,774A
 ; EARLIER FILING DATE: 1997-09-19
 ; EARLIER APPLICATION NUMBER: US 08/623,679
 ; EARLIER FILING DATE: 1996-03-29
 ; EARLIER APPLICATION NUMBER: US 08/412,431
 ; EARLIER FILING DATE: 1995-03-29
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 9
 ; LENGTH: 1533
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-08-933-774-9

Query Match 50.0%; Score 39; DB 3; Length 1533;
 Best Local Similarity 85.7%; Pred. NO. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 MGGSYRC 14
 DB 471 MGGAYRC 477
 Search completed: March 4, 2003, 16:57:58
 Job time : 8.69491 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:53:25 ; Search time 5.22034 Seconds
(without alignments)
113.093 Million cell updates/sec

Title: US-09-745-008-14
Perfect score: 78
Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	14	10	US-09-745-008-14
2	78	100.0	21	10	US-09-745-008-13
3	78	100.0	1162	10	US-09-745-008-34
4	67.5	86.5	45	10	US-09-745-008-12
5	40	51.3	230	10	US-09-741-669-372
6	39	50.0	756	10	US-09-946-175-2
7	39	50.0	761	10	US-09-946-175-3
8	39	50.0	1533	9	US-09-989-920-259
9	39	50.0	1533	10	US-09-828-466-7
10	39	50.0	1533	10	US-09-946-175-1
11	39	50.0	1863	9	US-09-832-292-29
12	39	50.0	1864	9	US-09-832-292-27
13	39	50.0	1885	10	US-09-920-346-2
14	38	48.7	453	9	US-10-028-072-270
15	38	48.7	453	9	US-10-121-049-270
16	38	48.7	453	9	US-10-123-904-270
17	38	48.7	453	9	US-10-140-470-270
18	38	48.7	453	9	US-10-175-746-270
19	38	48.7	453	9	US-10-176-918-270

20	38	48.7	453	9	US-10-176-921-270	Sequence 270, App
21	38	48.7	453	9	US-10-137-865-270	Sequence 270, App
22	38	48.7	453	9	US-10-140-474-270	Sequence 270, App
23	38	48.7	453	9	US-10-142-431-270	Sequence 270, App
24	38	48.7	453	9	US-10-143-114-270	Sequence 270, App
25	38	48.7	453	9	US-10-140-002-270	Sequence 270, App
26	38	48.7	1403	9	US-10-108-603-93	Sequence 93, Appl
27	37	47.4	146	10	US-09-923-246-56	Sequence 56, Appl
28	37	47.4	146	10	US-09-825-561A-47	Sequence 47, Appl
29	37	47.4	200	10	US-09-815-242-12759	Sequence 12759, A
30	37	47.4	216	10	US-09-815-242-5289	Sequence 5289, Ap
31	37	47.4	220	10	US-09-815-242-12235	Sequence 12235, A
32	37	47.4	400	9	US-10-078-770-114	Sequence 114, App
33	37	47.4	510	10	US-09-923-246-89	Sequence 89, Appl
34	37	47.4	816	9	US-09-796-720B-2	Sequence 2, Appli
35	36.5	46.8	235	9	US-10-017-910-8	Sequence 8, Appli
36	36.5	46.8	363	10	US-09-764-853-485	Sequence 485, App
37	36.5	46.8	390	9	US-09-965-529-19	Sequence 19, Appl
38	36	46.2	79	9	US-10-002-344A-196	Sequence 196, App
39	36	46.2	294	9	US-09-738-626-6805	Sequence 6805, Ap
40	36	46.2	332	10	US-09-740-273-2	Sequence 2, Appli
41	36	46.2	410	9	US-09-870-759-48	Sequence 48, Appl
42	35.5	45.5	560	10	US-09-730-525-18	Sequence 18, Appl
43	35.5	45.5	560	10	US-09-730-917-18	Sequence 18, Appl
44	35	44.9	87	10	US-09-864-761-40771	Sequence 40771, A
45	35	44.9	87	10	US-09-864-761-47374	Sequence 47374, A

ALIGNMENTS

RESULT 1
US-09-745-008-14
; Sequence 14, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; PRIOR APPLICATION NUMBER: 2000-12-20
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-14

Query Match 100.0%; Score 78; DB 10; Length 14;
Best Local Similarity 100.0%; Pred No. 1.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14
| | | | | | | | | | | | | |
Db 1 RQRLPKRMGGSYRC 14

RESULT 2
US-09-745-008-13
; Sequence 13, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor

```
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-13

Query Match      100.0%; Score 78; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYRC 14
Db 1 RORLPKRMGGSYRC 14

RESULT 3
US-09-745-008-34
; Sequence 34, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-34

Query Match      100.0%; Score 78; DB 10; Length 1162;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYRC 14
Db 379 RORLPKRMGGSYRC 392

RESULT 4
US-09-745-008-12
; Sequence 12, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-12

Query Match      100.0%; Score 78; DB 10; Length 1162;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RORLPKRMGGSYRC 14
Db 379 RORLPKRMGGSYRC 392

RESULT 5
US-09-741-669-372
; Sequence 372, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: Proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 372
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-372

Query Match      51.3%; Score 40; DB 10; Length 230;
Best Local Similarity 70.0%; Pred. No. 26;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RORLPKRMGG 10
Db 207 KQRIKEMGG 216

RESULT 6
US-09-946-175-2
; Sequence 2, Application US/09946175
; Patent No. US20020106671A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert
; TITLE OF INVENTION: IDENTIFICATION OF A CAPACITATIVE CALCIUM
; TITLE OF INVENTION: CHANNEL IN ANTIGEN PRESENTING CELLS AND USES THEREOF
; FILE REFERENCE: M1sn1(310800)
; CURRENT APPLICATION NUMBER: US/09/946,175
; CURRENT FILING DATE: 2001-09-05
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-175-2

Query Match      50.0%; Score 39; DB 10; Length 756;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 MGGSYRC 14
Db 471 MGGAYRC 477
```



```

; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 09/544,797
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1533
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-828-466-7

Query Match      50.0%; Score 39; DB 10; Length 1533;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGGSYRC 14
Db      471 MGGAYRC 477

RESULT 10
US-09-946-175-1
; Sequence 1, Application US/09946175
; Patent No. US20020106671A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert
; TITLE OF INVENTION: IDENTIFICATION OF A CAPACITATIVE CALCIUM
; TITLE OF INVENTION: CHANNEL IN ANTIGEN PRESENTING CELLS AND USES THEREOF
; FILE REFERENCE: Mlsn1(310800)
; CURRENT APPLICATION NUMBER: US/09/946,175
; CURRENT FILING DATE: 2001-09-05
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1533
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-175-1

Query Match      50.0%; Score 39; DB 10; Length 1533;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGGSYRC 14
Db      471 MGGAYRC 477

RESULT 11
US-09-832-292-29
; Sequence 29, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 601-1-098CIP
; CURRENT APPLICATION NUMBER: US/09/832,292
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 09/632,131
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 1863
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-832-292-29

Query Match      50.0%; Score 39; DB 9; Length 1863;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGGSYRC 14
Db      471 MGGAYRC 477

US-09-946-175-3
; Sequence 3, Application US/09946175
; Patent No. US20020106671A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert
; TITLE OF INVENTION: IDENTIFICATION OF A CAPACITATIVE CALCIUM
; TITLE OF INVENTION: CHANNEL IN ANTIGEN PRESENTING CELLS AND USES THEREOF
; FILE REFERENCE: Mlsn1(310800)
; CURRENT APPLICATION NUMBER: US/09/946,175
; CURRENT FILING DATE: 2001-09-05
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 761
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-946-175-3

Query Match      50.0%; Score 39; DB 10; Length 761;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGGSYRC 14
Db      471 MGGAYRC 477

RESULT 8
US-09-989-920-259
; Sequence 259, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 259
; LENGTH: 1533
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-259

Query Match      50.0%; Score 39; DB 9; Length 1533;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGGSYRC 14
Db      471 MGGAYRC 477

RESULT 9
US-09-828-466-7
; Sequence 7, Application US/09828466
; Patent No. US20020035056A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Immaculada
; TITLE OF INVENTION: 54420, A NOVEL HUMAN CALCIUM CHANNEL
; FILE REFERENCE: MNI-125CP
; CURRENT APPLICATION NUMBER: US/09/828,466
```

```
QY      8 MGSYRC 14
      |||:||||
Db      520 MGTYRC 526

RESULT 12
US-09-832-292-27
; Sequence 27, Application US/09832292
; Patent No. US20020177205A1
; GENERAL INFORMATION:
; APPLICANT: Ryazanov, Alexey
; TITLE OF INVENTION: MAMMALIAN ALPHA-KINASE PROTEINS, NUCLEIC ACIDS AND
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: 601-J-098CIP
; CURRENT APPLICATION NUMBER: US/09/832,292
; CURRENT FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 09/632,131
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 1864
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-832-292-27

Query Match      50.0%; Score 39; DB 9; Length 1864;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGSYRC 14
      |||:||||
Db      520 MGTYRC 526

RESULT 13
US-09-920-346-2
; Sequence 2, Application US/09920346
; Patent No. US2002081658A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: 18610, A NOVEL HUMAN TRANSIENT RECEPTOR
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: WNI-182
; CURRENT APPLICATION NUMBER: US/09/920,346
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: USSN 60/221,925
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1885
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-920-346-2

Query Match      50.0%; Score 39; DB 10; Length 1885;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 MGSYRC 14
      |||:||||
Db      520 MGTYRC 526

RESULT 14
US-10-028-072-270
; Sequence 270, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
```

; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064248
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/064809
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065846
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066453
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069212
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069278
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069334
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069694
; PRIOR FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: 60/072320
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 60/073612
; PRIOR FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: 60/074086
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/074092
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081818
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085149
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13

; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Query Match 48.7%; Score 38; DB 9; Length 453;

Best Local Similarity 77.8%; Pred. No. 1.1e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGGG 11

||||| |;

Db 43 RLPKRMGA 51

RESULT 15

US-10-121-049-270

; Sequence 270, Application US/10121049

; Publication No. US2003002239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: F330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
CURRENT FILING DATE: 2002-04-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 270
LENGTH: 453
TYPE: PRT
ORGANISM: Homo Sapien
US-10-121-049-270

Query Match 48.7%; Score 38; DB 9; Length 453;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLPKRMGS 11
Db 43 RLPKRMGA 51

Search completed: March 4, 2003, 16:58:00
Job time : 7.22034 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:53:45 ; Search time 31.5593 Seconds
(without alignments)
59.111 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 14
Sequence: 1 RQRLPKRMGGSYRC 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	42.9	62	AAU62077	Propionibacterium
2	6	42.9	73	ABP06708	Human ORFX protein
3	6	42.9	205	AAU65856	Propionibacterium
4	6	42.9	252	AAU96711	Human osteoclast a
5	6	42.9	263	AAU69232	Fcr-II protein seq
6	6	42.9	263	AAU96709	Human osteoclast a
7	6	42.9	271	AAU96713	Human osteoclast a
8	6	42.9	275	AAU96715	Human colon cancer
9	6	42.9	282	AAU96712	Human osteoclast a
10	6	42.9	329	ABG23911	Novel human diagno

11	6	42.9	332	22	AAE04313	Maize NPRI-interac
12	6	42.9	447	17	AAW01622	Oyster pearl prote
13	6	42.9	453	22	AAU12306	Human PRO197 poly
14	6	42.9	453	22	AAU53070	Human angiogenesis
15	6	42.9	453	23	AAU86128	Human PRO197 poly
16	6	42.9	456	22	ABE11591	Human angiopoietin
17	6	42.9	457	22	AAW25755	Human protein sequ
18	6	42.9	457	22	AAW25758	Human protein sequ
19	6	42.9	676	22	ABE63098	Drosophila melanog
20	6	42.9	957	22	AAW79016	Human protein SEQ
21	6	42.9	992	22	ABE65645	Drosophila melanog
22	5	35.7	9	23	ABG60440	Selective targetin
23	5	35.7	13	19	AAW66370	Indolicidin analog
24	5	35.7	13	21	AAW1853	Amino acid sequenc
25	5	35.7	15	21	AAW52191	MyoD derived pepti
26	5	35.7	21	8	AAW71703	Internal fragment
27	5	35.7	35	22	AAE61441	Human TANGO 275 EG
28	5	35.7	36	22	AAE61434	Human TANGO 275 EG
29	5	35.7	36	22	AAE61445	Human TANGO 275 EG
30	5	35.7	40	21	AAE20690	Polymetric immunogl
31	5	35.7	43	22	ABE39504	Peptide #7010 enco
32	5	35.7	43	22	ABE24254	Protein #6253 enco
33	5	35.7	43	22	AAE60196	Human brain expres
34	5	35.7	43	22	AAW72816	Human bone marrow
35	5	35.7	43	22	AAW19749	Peptide #6183 enco
36	5	35.7	43	22	AAE33046	Peptide #7083 enco
37	5	35.7	43	23	ABG42644	Human peptide enco
38	5	35.7	45	21	AAV59061	Tek receptor Notch
39	5	35.7	50	23	ABP34215	Human ORF3188 prot
40	5	35.7	50	23	ABP02762	Human ORFX protein
41	5	35.7	51	22	ABG26230	Novel human diagno
42	5	35.7	54	22	AAO13420	Human polypeptide
43	5	35.7	55	22	AAU56213	Propionibacterium
44	5	35.7	55	22	ABG30085	Novel human diagno
45	5	35.7	58	20	AAV17260	HLH domain of MyoD

ALIGNMENTS

RESULT 1
AAU62077
ID AAU62077 standard; Protein; 62 AA.
XX
AC AAU62077;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #22973.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.

N-PSDB; AAS59623.

Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris -

Example 1; SEQ ID No 23272; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hystertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

WPI; 2002-106308/14.
N-PSDB; ABN22460.

Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders -

Disclosure; SEQ ID 13398; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see table 1 in the specification). ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct sequences.

XX PA (CORI-) CORIXA CORP.
 XX SKKEY YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI; 2001-616774/71.
 DR N-PSDB; AAS59695.
 XX Protonibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX Example 1; SEQ ID No 27051; 1069pp; English.
 XX Sequences AAU39105-AAU68017 represent Protonibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 205 AA;

Query Match 42.9%; Score 6; DB 22; Length 205;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPK 6
 Db 102 RQRLPK 107

RESULT 4
 AAU96711
 ID AAU96711 standard; Protein; 252 AA.

XX AC AAU96711;

XX DT 30-JUL-2002 (first entry)

XX DE Human osteoclast associated receptor isoform C10.

XX KW Osteoclast associated; receptor; OSCAR; bone growth related disorder;
 KW bone resorption; osteoporosis; osteoporosis; immunotoxin;
 KW abnormal growth; development; repair; resorption; degradation;
 KW homeostasis of bone tissue; Paget's disease; osteogenesis imperfecta;
 KW fibrous dysplasia; hypophosphatasia; primary hyperparathyroidism;
 KW arthritis; periodontal disease; osteolysis; human; isoform C10.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..13

FT /label= Signal_peptide

FT Domain 14..112

FT /label= Immunoglobulin-like_domain

FT Protein 14..252

FT /label= Mature_OSCAR_isoform_C10

FT Domain 113..218
 FT /label= Immunoglobulin-like_domain
 FT Domain 219..237
 FT /label= Transmembrane_domain
 FT Domain 238..252
 FT /label= Cytoplasmic_tail_domain
 XX WO200220718-A2.

PN 14-MAR-2002.

XX 04-SEP-2001; 2001WO-US27502.

XX 05-SEP-2000; 2000US-230152P.

PR 24-JUL-2001; 2001US-307557P.

XX (UYRO) UNIV ROCKEFELLER.

PA (UYPE-) UNIV PENNSYLVANIA.

XX Choi Y, Kim N;

XX WPI; 2002-362244/39.

DR N-PSDB; ABK50892.

XX New osteoclast associated receptor polypeptide that modulates activity
 PT of osteoclast cells, involved in growth, development, repair and
 PT homeostasis of bone tissue and for treatment of osteoporosis and
 PT osteoporosis -
 XX Claim 9; Fig 5B; 160pp; English.

XX The invention describes a novel isolated osteoclast associated receptor
 CC (OSCAR) polypeptide (I). An OSCAR gene, polypeptides and antibodies
 CC directed against the polypeptide are useful for both prognostic and
 CC diagnostic applications for evaluating bone growth related disorders,
 CC and identifying subjects having a predisposition to the disorders.
 CC For example, a compound that increases the activity of an OSCAR cell
 CC product is useful for increasing the activity of an osteoclast cell for
 CC increasing bone resorption and treating a bone growth related disorder,
 CC especially osteoporosis, and a compound that decreases the activity of
 CC an OSCAR gene product is useful for decreasing activity of an osteoclast
 CC cell for decreasing bone resorption, for treating osteoporosis. An
 CC antibody to (I) is useful for diagnostics and intracellular regulation of
 CC OSCAR activity, to isolate cells which express OSCAR polypeptide and to
 CC create immunotoxins. (I) can be used in a treatment for disorders
 CC involving abnormal growth, development, repair, resorption, degradation,
 CC or homeostasis of bone tissue, such as Paget's disease, osteogenesis
 CC imperfecta, fibrous dysplasia, hypophosphatasia, primary
 CC hyperparathyroidism, arthritis, periodontal disease and osteolysis.
 CC This is the amino acid sequence of the human osteoclast associated
 CC receptor (OSCAR) isoform C10.

XX SQ Sequence 252 AA;

Query Match 42.9%; Score 6; DB 23; Length 252;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 GGSYRC 14

Db 84 GGSYRC 89

RESULT 5

AAW69232

XX ID AAW69232 standard; Protein; 263 AA.

XX AC AAW69232;

XX DT 20-OCT-1998 (first entry)

XX FCRI-II protein sequence.

KW Fc receptor-like protein; phagocytosis inducer; rheumatoid arthritis;
 KW immune complex related disease; systemic lupus erythematosus; allergy;
 KW haemolytic anaemia; thrombocytopaenia; anaphylaxis; cancer; lymphoma;
 KW leukaemia; infection; immunomodulator; viral entry inhibitor; therapy;
 KW FcR-II.
 XX
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..18 /note= "signal peptide"
 FT Protein 19..263 /note= "mature FcR-II"
 FT
 FT
 XX WO9831806-A2.
 XX
 XX 23-JUL-1998.
 XX
 XX 20-JAN-1998; 98WO-US01184.
 XX
 XX 18-JUN-1997; 97US-0049872.
 PR 21-JAN-1997; 97US-0034205.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Gentz RL, Murphy M, Ni J, Olsen HS, Ruben SM;
 XX WPI; 1998-414105/35.
 DR N-PSDB; AAV44825.
 XX
 XX Nucleic acid encoding Fc receptor-like polypeptides or their
 PT fragments - and related vectors, transformed cells and antibodies,
 PT useful for treating and diagnosing diseases of the haematopoietic
 PT and immune systems
 XX
 XX Claim 23; Fig 2A; 141pp; English.
 PS
 XX This sequence is the Fc receptor-like II protein (FcR-II) of the
 CC invention. Cells containing the DNA are used to express the recombinant
 CC protein, and to screen for specific (ant)agonists. The proteins are used
 CC to induce phagocytosis, and their (ant)agonists are used to treat immune
 CC complex related diseases (e.g. rheumatoid arthritis, systemic lupus
 CC erythematosus, haemolytic anaemia, thrombocytopaenia, anaphylaxis,
 CC allergy, colorectal or breast cancer, lymphoma, leukaemia, infection by
 CC intracellular pathogens etc). The antagonists are also useful as
 CC immunomodulators and inhibitors of viral (e.g. human immune deficiency or
 CC dengue viruses) entry into cells. The proteins may also be used to screen
 CC for specific binding agents, i.e. (ant)agonists, for raising antibodies
 CC (Ab), and for identification of particular cells or tissues. The Ab can
 CC be used therapeutically as antagonists; as assay reagents for diagnostic
 CC determination of the levels of expression of the proteins and for
 CC affinity purification of the proteins. The DNA and its fragments are
 CC useful as hybridisation probes or primers for isolating related genes, in
 CC situ hybridisation (chromosome mapping) and diagnostically to measure
 CC mRNA expression.
 XX
 XX Query Match 42.9%; Score 6; DB 19; Length 263;
 XX Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 9 GGSYRC 14
 Db 95 GGSYRC 100
 RESULT 6
 AAU96709
 ID AAU96709 standard; Protein; 263 AA.
 XX
 XX AAU96709;
 XX

DT 30-JUL-2002 (first entry)
 XX Human osteoclast associated receptor isoform C18.
 DE
 XX Osteoclast associated; receptor; OSCAR; bone growth related disorder;
 KW bone resorption; osteopetrosis; osteoporosis; immunotoxin;
 KW abnormal growth; development; repair; resorption; degradation;
 KW homeostasis of bone tissue; Paget's disease; osteogenesis imperfecta;
 KW fibrous dysplasia; hypophosphatasia; primary hyperparathyroidism;
 KW arthritis; periodontal disease; osteolysis; human; isoform C18.
 XX
 OS Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 39..18 /label= Signal peptide
 FT Domain 19..123 /label= Immunoglobulin-like_domain
 FT Protein 19..263 /label= Mature_OSCAR_isoform_C18
 FT Domain 124..229 /label= Immunoglobulin-like domain
 FT Domain 230..248 /label= Transmembrane_domain
 FT Domain 249..263 /label= Cytoplasmic_tail_domain
 FT
 FT
 XX WO200220718-A2.
 PN
 XX 14-MAR-2002.
 XX
 XX 04-SEP-2001; 2001WO-US27502.
 PF
 XX 05-SEP-2000; 2000US-230152P.
 PR 24-JUL-2001; 2001US-307557P.
 PR
 XX (UYRO) UNIV ROCKEFELLER.
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA
 XX Choi Y, Kim N;
 FI
 XX WPI; 2002-362244/39.
 XX N-PSDB; ABK50890.
 DR
 XX New osteoclast associated receptor polypeptide that modulates activity
 PT of osteoclast cells, involved in growth, development, repair and
 PT homeostasis of bone tissue and for treatment of osteoporosis and
 PT osteopetrosis -
 XX
 PS Claim 7; Fig 3B; 160pp; English.
 XX
 CC The invention describes a novel isolated osteoclast associated receptor
 CC (OSCAR) polypeptide (I). An OSCAR gene, polypeptides and antibodies
 CC directed against the polypeptide are useful for both prognostic and
 CC diagnostic applications for evaluating bone growth related disorders,
 CC and identifying subjects having a predisposition to the disorders.
 CC For example, a compound that increases the activity of an OSCAR gene
 CC product is useful for increasing the activity of an osteoclast cell for
 CC increasing bone resorption and treating a bone growth related disorder,
 CC especially osteopetrosis, and a compound that decreases the activity of
 CC an OSCAR gene product is useful for decreasing activity of an osteoclast
 CC cell for decreasing bone resorption, for treating osteoporosis. An
 CC antibody to (I) is useful for diagnostics and intracellular regulation of
 CC OSCAR activity, to isolate cells which express OSCAR polypeptide and to
 CC create immunotoxins. (I) can be used in a treatment for disorders
 CC involving abnormal growth, development, repair, resorption, degradation,
 CC or homeostasis of bone tissue, such as Paget's disease, osteogenesis
 CC imperfecta, fibrous dysplasia, hypophosphatasia, primary
 CC hyperparathyroidism, arthritis, periodontal disease and osteolysis.
 CC This is the amino acid sequence of the human osteoclast associated
 CC receptor (OSCAR) isoform C18.
 XX
 XX Sequence 263 AA;


```

Query Match      42.9%; Score 6; DB 23; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYRC 14
   |||||
Db 95 GGSYRC 100

RESULT 7
AAU96713
ID AAU96713 standard; Protein; 271 AA.
XX
AC AAU96713;
DT 30-JUL-2002 (first entry)
XX
DE Human osteoclast associated receptor splice variant S2.
XX
KW Osteoclast associated; receptor; OSCAR; bone growth related disorder;
KW bone resorption; osteopetrosis; osteoporosis; immunotoxin;
KW abnormal growth; development; repair; resorption; degradation;
KW homeostasis of bone tissue; Paget's disease; osteogenesis imperfecta;
KW fibrous dysplasia; hypophosphatasia; primary hyperparathyroidism;
KW arthritis; periodontal disease; osteolysis; human; splice variant S2.
XX
OS Homo sapiens.
XX
PN WO200220718-A2.
XX
PD 14-MAR-2002.
XX
PF 04-SEP-2001; 2001WO-US27502.
XX
PR 05-SEP-2000; 2000US-230152P.
PR 24-JUL-2001; 2001US-307557P.
XX
PA (UYRQ) UNIV ROCKEFELLER.
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Choi Y, Kim N;
XX
WPI; 2002-362244/39.
DR N-PSDB; ABK50907.
XX
PT New osteoclast associated receptor polypeptide that modulates activity
PT of osteoclast cells, involved in growth, development, repair and
PT homeostasis of bone tissue and for treatment of osteoporosis and
PT osteopetrosis -
XX
Claim 104; Fig 25B; 160pp; English.
XX
PS
XX
CC The invention describes a novel isolated osteoclast associated receptor
CC (OSCAR) polypeptide (I). An OSCAR gene, polypeptides and antibodies
CC directed against the polypeptide are useful for both prognostic and
CC diagnostic applications for evaluating bone growth related disorders,
CC and identifying subjects having a predisposition to the disorders.
CC For example, a compound that increases the activity of an OSCAR gene
CC product is useful for increasing the activity of an osteoclast cell for
CC increasing bone resorption and treating a bone growth related disorder,
CC especially osteopetrosis, and a compound that decreases the activity of
CC an OSCAR gene product is useful for decreasing activity of an osteoclast
CC cell for decreasing bone resorption, for treating osteoporosis. An
CC antibody to (I) is useful for diagnostics and intracellular regulation of
CC OSCAR activity, to isolate cells which express OSCAR polypeptide and to
CC create immunotoxins. (I) can be used in a treatment for disorders
CC involving abnormal growth, development, repair, resorption, degradation,
CC or homeostasis of bone tissue, such as Paget's disease, osteogenesis
CC imperfecta, fibrous dysplasia, hypophosphatasia, primary
CC hyperparathyroidism, arthritis, periodontal disease and osteolysis.
CC This is the amino acid sequence of the human osteoclast associated
CC receptor (OSCAR) splice variant S2.

```

```

XX SQ Sequence 271 AA;
Query Match      42.9%; Score 6; DB 23; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYRC 14
   |||||
Db 84 GGSYRC 89

RESULT 8
AAG75615
ID AAG75615 standard; Protein; 275 AA.
XX
AC AAG75615;
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6379.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI; 2001-235357/24.
DR N-PSDB; AAH35020.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
Claim 11; Page 7854-7856; 9803pp; English.
XX
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing P.
CC Inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 275 AA;
Query Match      42.9%; Score 6; DB 22; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 9 GGSYRC 14
 Db 107 GGSYRC 112

RESULT 9
 AAU96712
 ID AAU96712 standard; Protein; 282 AA.
 AC AAU96712;
 DT 30-JUL-2002 (first entry)
 XX Human osteoclast associated receptor splice variant S1.
 DE
 XX Osteoclast associated; receptor; OSCAR; bone growth related disorder;
 KW bone resorption; osteoporosis; osteopenia; immunotoxin;
 KW abnormal growth; development; repair; resorption; degradation;
 KW homeostasis of bone tissue; Paget's disease; osteogenesis imperfecta;
 KW fibrous dysplasia; hypophosphatasia; primary hyperparathyroidism;
 KW arthritis; periodontal disease; osteolysis; human; splice variant S1.
 XX
 OS Homo sapiens.
 XX
 XX WO200220718-A2.
 PN 14-MAR-2002.
 XX
 PD 04-SEP-2001; 2001WO-US17502.
 XX
 PF 05-SEP-2000; 2000US-230152P.
 PR 24-JUL-2001; 2001US-307557P.
 XX
 XX (UYRQ) UNIV ROCKEFELLER.
 PA (UYPE-) UNIV PENNSYLVANIA.
 XX
 XX Choi Y, Kim N;
 XX WPI; 2002-362244/39.
 DR N-PSDB; ABK50906.
 XX
 XX New osteoclast associated receptor polypeptide that modulates activity
 PT of osteoclast cells, involved in growth, development, repair and
 PT homeostasis of bone tissue and for treatment of osteoporosis and
 PT osteopetrosis -
 XX
 PS Claim 103; Fig 24B; 160pp; English.
 XX
 CC The invention describes a novel isolated osteoclast associated receptor
 CC (OSCAR) polypeptide (I). An OSCAR gene, polypeptides and antibodies
 CC directed against the polypeptide are useful for both prognostic and
 CC diagnostic applications for evaluating bone growth related disorders,
 CC and identifying subjects having a predisposition to the disorders.
 CC For example, a compound that increases the activity of an OSCAR gene
 CC product is useful for increasing the activity of an osteoclast cell for
 CC increasing bone resorption and treating a bone growth related disorder,
 CC especially osteopetrosis, and a compound that decreases the activity of
 CC an OSCAR gene product is useful for decreasing activity of an osteoclast
 CC cell for decreasing bone resorption, for treating osteoporosis. An
 CC antibody to (I) is useful for diagnostics and intracellular regulation of
 CC OSCAR activity, to isolate cells which express OSCAR polypeptide and to
 CC create immunotoxins. (I) can be used in a treatment for disorders
 CC involving abnormal growth, development, repair, resorption, degradation,
 CC or homeostasis of bone tissue, such as Paget's disease, osteogenesis
 CC imperfecta, fibrous dysplasia, hypophosphatasia, primary
 CC hyperparathyroidism, arthritis, periodontal disease and osteolysis.
 CC This is the amino acid sequence of the human osteoclast associated
 CC receptor (OSCAR) splice variant S1.
 XX
 SQ Sequence 282 AA;
 Query Match 42.9%; Score 6; DB 23; Length 282;
 Best Local Similarity 100.0%; Pred. No. 1.3e-02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYRC 14
 Db 95 GGSYRC 100

RESULT 10
 ABG23911
 ID ABG23911 standard; Protein; 329 AA.
 AC ABG23911;
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #23902.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN 11-OCT-2001.
 XX
 PD 30-MAR-2001; 2001WO-US08631.
 XX
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS88098.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 54270; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 329 AA;
 Query Match 42.9%; Score 6; DB 22; Length 329;
 Best Local Similarity 100.0%; Pred. No. 1.3e-02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 3 RLPKRM 8
Db 35 RLPKRM 40

RESULT 11
AAE04313
ID AAE04313 standard; Protein; 332 AA.
AC AAE04313;
XX
DT 04-SEP-2001 (first entry)
DE
DE Maize NPR1-interacting protein #1.
XX
XX Maize; NPR1-interacting protein; disease resistance; sequence shuffling;
XX transgenic plant; signal transduction pathway.
XX
OS Zea mays.
XX
XX WO200146423-A2.
XX
XX 28-JUN-2001.
XX
XX 19-DEC-2000; 2000WO-US34524.
XX
XX 21-DEC-1999; 99US-0171691.
XX
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Crane EH;
XX
XX WPI; 2001-408649/43.
XX
XX N-PSDB; AAD08578.
XX
XX Novel maize NPR1-interacting polynucleotide, useful for engineering
XX plants with improved disease resistance by increasing sensitivity or
XX capacity of signal transduction pathway and for sequence shuffling -
XX
XX Claim 12; Page 65; 69pp; English.
XX
XX The invention relates to NPR1-interacting proteins and nucleic
XX acids encoding them. NPR1-interacting DNA is useful for modulating
XX the level of NPR1-interacting protein in plants such as maize,
XX soybean etc. By manipulating NPR1-interacting DNA in maize or in
XX other plants, the plant can be engineered to improve resistance to
XX pathogens by increasing the sensitivity or capacity of the signal
XX transduction pathway. The plants containing altered NPR1 expression
XX are useful as universal disease susceptible plants. NPR1-interacting
XX DNA is further useful for sequence shuffling. They are also used as
XX probes. The invention also provides transgenic plants with increased
XX disease resistance. The present sequence is NPR1-interacting protein.
XX
SQ Sequence 332 AA;
Query Match 42.9%; Score 6; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 QRLPKR 7
Db 216 QRLPKR 221

RESULT 12
AAW01622
ID AAW01622 standard; Protein; 447 AA.
XX
XX AAW01622;
XX
XX 05-AUG-1997 (first entry)
XX
XX

```

```

DE Oyster pearl protein, nacrein.
XX
XX Oyster; pearl; nacrein; mother of pearl; nacre; expression; vector;
XX recombinant; industrial; production.
XX
XX Pinctada fucata.
XX
XX Key Location/Qualifiers
XX Peptide 1..17
XX FT /label= sig_peptide
XX FT 18..447
XX FT /label= mat_peptide
XX FT Active-site 375..422
XX
XX PN WO9635786-A1.
XX
XX 14-NOV-1996.
XX
XX 09-MAY-1996; 96WO-JP01236.
XX
XX 09-MAY-1995; 95JP-0110877.
XX
XX (MATS/) MATSUSHIRO A.
XX (FUJI/) FUJIKAWA Y.
XX
XX Matsushiro A;
XX
XX WPI; 1996-518674/51.
XX
XX N-PSDB; AAT58313.
XX
XX Recombinant production of nacrein from Pinctada fucata - for use in
XX pearl production on industrial scale.
XX
XX Claim 4; Pages 20-24; 46pp; English.
XX
XX The present sequence is the oyster pearl protein, nacrein,
XX which has a molecular weight of about 60000. The nacrein cDNA can
XX be incorporated into an expression vector for the transformation of
XX a suitable host (e.g. E. coli, yeast or silkworm), useful in the
XX efficient, large scale production of nacrein for use in industrial
XX pearl production.
XX
XX mRNA was extracted from the nacreous layer of the oyster Pinctada
XX fucata and used to generate a lambda gt10 cDNA library, which was
XX screened using a synthetic probe. The DNA fragment encoding the
XX nacrein gene was incorporated into pT7-7 vector and used to
XX transform E. coli BL21 (lambda DE3). Culture of this transformant
XX produced nacrein.
XX
XX Sequence 447 AA;
Query Match 42.9%; Score 6; DB 17; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RQRLPK 6
Db 94 RQRLPK 99

RESULT 13
AAU12306
ID AAU12306 standard; Protein; 453 AA.
XX
XX AAU12306;
XX
XX 24-OCT-2001 (first entry)
XX
XX Human PRO197 polypeptide sequence.
XX
XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
XX breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
XX cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
XX adipocyte; A-peptide; factor VIIA; gene therapy.
XX

```

XX OS Homo sapiens.
 XX WO200140466-A2.
 XX 07-JUN-2001.
 XX 01-DEC-2000; 2000WO-US32678.
 XX 01-DEC-1999; 99WO-US28301.
 XX 01-DEC-1999; 99WO-US28634.
 XX 02-DEC-1999; 99WO-US28551.
 XX 02-DEC-1999; 99WO-US28564.
 XX 02-DEC-1999; 99WO-US28565.
 XX 09-DEC-1999; 99US-0170262.
 XX 16-DEC-1999; 99WO-US30095.
 XX 20-DEC-1999; 99WO-US30911.
 XX 20-DEC-1999; 99WO-US30999.
 XX 30-DEC-1999; 99WO-US31243.
 XX 06-JAN-2000; 2000WO-US00277.
 XX 11-FEB-2000; 2000WO-US00376.
 XX 18-FEB-2000; 2000WO-US04341.
 XX 18-FEB-2000; 2000WO-US04342.
 XX 24-FEB-2000; 2000WO-US04914.
 XX 24-FEB-2000; 2000WO-US05004.
 XX 01-MAR-2000; 2000WO-US05601.
 XX 20-MAR-2000; 2000WO-US07377.
 XX 21-MAR-2000; 2000WO-US07532.
 XX 30-MAR-2000; 2000WO-US08439.
 XX 17-MAY-2000; 2000WO-US13705.
 XX 22-MAY-2000; 2000WO-US14042.
 XX 30-MAY-2000; 2000WO-US14941.
 XX 02-JUN-2000; 2000WO-US15284.
 XX 10-NOV-2000; 2000WO-US30873.
 XX (GETH) GENENTECH INC.
 XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-408281/43.
 DR N-PSDB; AAS21378.
 XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical -
 XX Claim 12; Fig 270; 813pp; English.
 XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.

XX SQ Sequence 453 AA;
 Query Match 42.9%; Score 6; DB 22; Length 453;
 Best Local Similarity 100.0%; Pred. No. 1.7e-02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 RLPKRM 8
 Db 43 RLPKRM 48
 RESULT 14
 AAB53070
 ID AAB53070 standard; Protein; 453 AA.
 XX AC AAB53070;
 XX DT 28-FEB-2001 (first entry)
 XX Human angiogenesis-associated protein PRO197, SEQ ID NO:31.
 DE Human; angiogenesis-associated protein; PRO; endothelial cell growth;
 KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
 KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
 KW myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
 KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
 KW Alzheimer's disease; Huntington's disease; stroke; drug screening;
 KW gene therapy; transgenic animal.
 XX OS Homo sapiens.
 XX WO2000053753-A2.
 PN 14-SEP-2000.
 PD 05-JAN-2000; 2000WO-US00219.
 XX 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 14-MAY-1999; 99US-0134287.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 08-SEP-1999; 99WO-US20594.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-NOV-1999; 99WO-US28409.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 XX (GETH) GENENTECH INC.
 XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;
 PI Godowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA;
 PI Praeli NF, Pitti RM, Watanabe CK, Williams PM, Wood WI;
 XX WPI; 2001-090793/10.
 DR N-PSDB; AAC97388.
 XX New isolated nucleic acid for producing a PRO polypeptide, analyzing
 PT genetic disorders and treating cardiovascular, endothelial or
 PT angiogenic disorders, such as atherosclerosis, wounds or cancer -
 XX Claim 69; Fig 14; 293pp; English.
 XX The invention relates to novel human angiogenesis-associated proteins
 CC designated PRO proteins (AAB53084-B53097), and to nucleic acids encoding
 CC PRO proteins. The invention also relates to vectors and host cells

comprising a PRO nucleic acid, the recombinant production of a PRO protein, PRO antibodies specific for a PRO protein, fusion proteins comprising a PRO protein, agonists or antagonists of a PRO protein, and compounds which inhibit the expression of a PRO gene. The invention additionally encompasses methods of identifying modulators of PRO expression or activity; diagnosing a cardiovascular, endothelial or angiogenic disorder, or a susceptibility to such a disorder by detecting mutations in a PRO gene, or the expression level of a PRO gene within a particular tissue; treating a cardiovascular, endothelial or angiogenic disorder via the administration of a PRO protein, PRO nucleic acid, or PRO agonist or antagonist; a retroviral gene therapy vector comprising a PRO nucleic acid; and methods of inhibiting or stimulating endothelial cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the administration of a PRO protein, or an agonist or antagonist thereof. PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO agonists and PRO antagonists may be used as therapeutic agents to treat cardiovascular, endothelial or angiogenic disorders, such as atherosclerosis, osteoporosis, myocardial infarction, hypertension, diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis, endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's disease, or stroke. PRO nucleic acids are additionally useful in the recombinant production of PRO proteins, as hybridisation probes to screen libraries to isolate cDNAs with sequence identity to PRO proteins, to map genes encoding PRO proteins, to analyse genetic disorders, and in gene therapy. PRO nucleic acids can also be used to produce transgenic animals useful for the development and screening of potential therapeutic agents. The present sequence represents a PRO protein of the invention.

XX Sequence 453 AA;

Query Match 42.9%; Score 6; DB 22; Length 453;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8

Db 43 RLPKRM 48

RESULT 15

AAU86128

ID AAU86128 standard; Protein; 453 AA.

XX AC AAU86128;

XX AC AAU86128;

DT 15-JUL-2002 (first entry)

XX Human PRO197 polypeptide.

XX Human; PRO; benign tumour; malignant tumour; lymphoid malignancy; leukaemia; neuronal disorder; stromal disorder; blastocoelec disorder; inflammatory disorder; immune disorder; angiogenic disorder; cytostatic; neuroprotective.

XX OS Homo sapiens.

XX PN WO200153486-A1.

XX PD 26-JUL-2001.

XX PF 11-FEB-2000; 2000WO-US03565.

XX PR 08-MAR-1999; 99WO-US05028.

PR 11-MAR-1999; 99US-123972P.

PR 11-MAY-1999; 99US-133459P.

PR 02-JUN-1999; 99WO-US12252.

PR 22-JUN-1999; 99US-140650P.

PR 20-JUN-1999; 99US-140653P.

PR 20-JUL-1999; 99US-144758P.

PR 26-JUL-1999; 99US-145698P.

PR 28-JUL-1999; 99US-146222P.

PR 17-AUG-1999; 99US-149395P.

PR 31-AUG-1999; 99US-151689P.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 30-NOV-1999; 99WO-US28313.

PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 05-JAN-2000; 2000WO-US00219.

XX (GETH) GENENTECH INC.

PA Ashkenazi AJ, Goddard A, Godowski PJ, Gurney AL, Hillan KJ;

PI Marksters SA, Pan J, Pitti RM, Roy MA, Smith V, Stone DM;

PI Watanabe CK, Wood WI;

XX WPI; 2002-205567/26.

DR N-PSDB; ABK40254.

XX Thirty five nucleic acids encoding PRO polypeptides; useful for

PT treating benign or malignant tumours, leukaemias and lymphoid

PT malignancies, inflammatory, angiogenic and immunologic disorders -

XX Claim 61; Fig 2; 302pp; English.

XX The present invention relates to the isolation of novel human PRO

CC polypeptides and the polynucleotide sequences encoding them. The

CC PRO polypeptides, agonists, antagonists or anti-PRO antibodies are

CC useful for treating benign or malignant tumours (e.g. renal, kidney,

CC bladder, breast, etc), leukaemias and lymphoid malignancies, other

CC disorders such as neuronal, glial, astrocytal, hypothalamic, glandular,

CC macrophagal, stromal and blastocoelec disorders, inflammatory, immune

CC and angiogenic disorders. The polynucleotide sequences are also

CC useful in gene therapy. AAU86128-AAU86162 represent the human PRO

CC polypeptides of the invention.

XX Sequence 453 AA;

Query Match 42.9%; Score 6; DB 23; Length 453;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8

Db 43 RLPKRM 48

Search completed: March 4, 2003, 17:01:04

Job time : 87.5593 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Query Match 42.9%; Score 6; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KRMGS 11
Db 8 KRMGS 13

RESULT 3
R5BSL5
ribosomal protein L15 - *Bacillus subtilis*
C;Species: *Bacillus subtilis*
C;Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text_change 16-Jun-2000
C;Accession: S12682; JS0491; S08628; A69696
R;Yoshikawa, H.; Doi, R.H.
Nucleic Acids Res. 18, 1647, 1990
A;Title: Sequence of the *Bacillus subtilis* spectinomycin resistance gene region.
A;Reference number: S12680; MUID:90221911; PMID:21139212
A;Accession: S12682
A;Molecule type: DNA
A;Residues: 1-146 <YOS>
A;Cross-references: EMBL:M31102; NID:g1184272; PIDN:AAB59117.1; PID:g143577
R;Nakamura, K.; Nakamura, A.; Takamatsu, H.; Yoshikawa, H.; Yamane, K.
J. Biochem. 107, 603-607, 1990
A;Title: Cloning and characterization of a *Bacillus subtilis* gene homologous to *E. coli*
A;Reference number: JS0490; MUID:90292990; PMID:2113521
A;Accession: JS0491
A;Molecule type: DNA
A;Residues: 1-146 <NAK>
A;Cross-references: DDBJ:D00619; NID:g216336; PIDN:BAA00494.1; PID:g216338
R;Suh, J.W.; Boylan, S.A.; Thomas, S.M.; Dolan, K.M.; Oliver, D.B.; Price, C.W.
Mol. Microbiol. 4, 305-314, 1990
A;Title: Isolation of a secY homologue from *Bacillus subtilis*: evidence for a common pro
A;Reference number: S08628; MUID:90251170; PMID:2110998
A;Accession: S08628
A;Molecule type: DNA
A;Residues: 94-146 <SUH>
A;Cross-references: EMBL:X51329; NID:g40132; PIDN:CAA35711.1; PID:g40133
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Patro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69596
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-146 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11911.1; PID:g2632402
A;Experimental source: strain 168
C;Genetics:
A;Gene: rplO
C;Superfamily: *Escherichia coli* ribosomal protein L15
C;Keywords: protein biosynthesis; ribosome

Query Match 42.9%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRLPKR 7
Db 59 QRLPKR 64

RESULT 4
S70810
type IV prepilin peptidase (EC 3.4.99.-) xpsO - *Xanthomonas campestris* pv. *campestris*
N;Contains: type IV pilin N-methyltransferase (EC 2.1.1.-)
C;Species: *Xanthomonas campestris* pv. *campestris*
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S70810
R;Hu, N.T.; Lee, P.F.; Chen, C.
Mol. Microbiol. 18, 769-777, 1995
A;Title: The type IV pre-pilin leader peptidase of *Xanthomonas campestris* pv. *campestris*
A;Reference number: S70809; MUID:96414476; PMID:8817497
A;Accession: S70810
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-287 <HUN>
A;Cross-references: EMBL:U12432; NID:g529681; PIDN:AAC43571.1; PID:g529683
C;Genetics:
A;Gene: xpsO
C;Superfamily: type IV prepilin peptidase
C;Keywords: hydrolase; methyltransferase; S-adenosylmethionine

Query Match 42.9%; Score 6; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPKRM 8
Db 32 RLPKRM 37

RESULT 5
T24650
hypothetical protein T07C5.2 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T24650
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19917
A;Accession: T24650
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-334 <WIL>
A;Cross-references: EMBL:Z50006; PIDN:CAA90299.1; GSPDB:GN00028; CESP:T07C5.2
A;Experimental source: clone T07C5
C;Genetics:
A;Gene: CESP:T07C5.2
A;Map position: X
A;Introns: 20/3; 250/1

Query Match 42.9%; Score 6; DB 2; Length 334;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQLPKK 6
Db 57 RQLPKK 62

RESULT 6
A47312
NS34 homolog - murine rotavirus B (strain IDIR)
N;Alternate names: gene 5 protein
C;Species: group B rotavirus
A;Note: strain IDIR (infectious diarrhea of infant rats)
C;Date: 28-Mar-1994 #sequence revision 25-Apr-1997 #text_change 18-Jul-2001
C;Accession: A47312
R;Eiden, J.J.
Virology 196, 298-302, 1993
A;Title: Gene 5 of the IDIR agent (group B rotavirus) encodes a protein equivalent to NS:
A;Reference number: A47312; MUID:93362418; PMID:8395119

A;Accession: A47312
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: genomic RNA
A;Residues: 1-390 <EID>
A;Cross-references: GB:L09722; NID:G397809
A;Note: sequence extracted from NCBI backbone (NCBIP:136603)
A;Note: this translation is not annotated in GenBank entry ROTNS34A, release 113.0

Query Match 42.9%; Score 6; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPRM 8
|||||
Db 51 RLPRM 56

RESULT 7
T04434
replication protein A1 homolog T18B16.100 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 02-Sep-2000
A;Accession: T04434
R;Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A;Reference number: Z15359
A;Accession: T04434
A;Molecule type: DNA
A;Residues: 1-717 <BEV>
A;Cross-references: EMBL:AL021687
A;Experimental source: cultivar Columbia; BAC clone T18B16
C;Genetics:
A;Map position: 4
A;Introns: 96/2; 165/1; 350/3
A;Note: T18B16.100
C;Superfamily: replication protein A1

Query Match 42.9%; Score 6; DB 2; Length 717;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYR 13
|||||
Db 694 MGGSYR 699

RESULT 8
G83637
hypothetical protein PA0077 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
A;Accession: G83637
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miroguchi, S.D.; Warriner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: G83637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1101 <STO>
A;Cross-references: GB:AE004446; GB:AE004091; NID:g9945886; PIDN:AAG03467.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0077

Query Match 42.9%; Score 6; DB 2; Length 1101;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQLPK 6
|||||

Db 583 RQLPK 588

RESULT 9
SYBYVT
valine-tRNA ligase (EC 6.1.1.9) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G4647; protein YGR094w; valyl-tRNA synthetase
C;Species: Saccharomyces cerevisiae
C;Date: 30-Jun-1992 #sequence_revision 19-Jul-1996 #text_change 03-Jun-2002
A;Accession: S64389; S64399; A29871
R;Wedler, H.; Scharfe, M.; Wedler, E.; Wambutt, R.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64356
A;Accession: S64389
A;Molecule type: DNA
A;Residues: 1-1104 <WED>
A;Cross-references: EMBL:Z72879; NID:gl323140; PIDN:CAA97097.1; PID:gl323141; GSPDB:GN001
A;Experimental source: strain S288C
R;Hernandez, K.; Weber, N.; Wipfli, P.; Schmidheini, T.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64392
A;Accession: S64399
A;Molecule type: DNA
A;Residues: 1-1104 <HER>
A;Cross-references: EMBL:Z72879; NID:gl323140; PIDN:CAA97097.1; PID:gl323141; GSPDB:GN001
A;Experimental source: strain S288C
R;Jordana, X.; Chatton, B.; Paz-Weishaar, M.; Buhler, J.M.; Cramer, F.; Ebel, J.P.; Fasi
J. Biol. Chem. 262, 7189-7194, 1987
A;Title: Structure of the yeast valyl-tRNA synthetase gene (VASI) and the homology of its
A;Reference number: A29871; MUID:8722321; PMID:3294828
A;Accession: A29871
A;Molecule type: DNA
A;Residues: 1-146, 'G', 148-539, 'K', 541-1104 <JOR>
A;Cross-references: GB:J02719; NID:gl73163; PIDN:AAA35207.1; PID:gl73164
C;Genetics:
A;Gene: SGD:VAS1; MIPS:YGR094w
A;Cross-references: SGD:S0003326; MIPS:YGR094w
A;Map position: 7R
C;Superfamily: valine-tRNA ligase
C;Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 42.9%; Score 6; DB 1; Length 1104;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RQLPK 7
|||||
Db 913 RQLPK 918

RESULT 10
S47383
T-cell antigen receptor VJ junction beta chain - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 05-Nov-1999
A;Accession: S47383
R;Lehner, P.J.
submitted to the EMBL Data Library, August 1994
A;Description: Human HLA-A0201 restricted recognition of influenza A is dominated by T ce
A;Reference number: S47355
A;Accession: S47383
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-13 <LEH>
A;Cross-references: EMBL:Z35709; NID:g527513; PIDN:CAA84778.1; PID:g527514
C;Keywords: T-cell receptor

Query Match 35.7%; Score 5; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYR 12
|||||

Db 5 MGGSY 9

RESULT 11

D58213
protamine III - American alligator
C;Species: Alligator mississippiensis (American alligator)
C;Date: 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 07-May-1999
C;Accession: D58213
R;Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.
J. Biol. Chem. 271, 23547-23557, 1996
A;Title: Protamines of reptiles.
A;Reference number: A58208; MUID:96394458; PMID:8798564
A;Accession: D58213
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-43 <HUN>
C;Superfamily: sperm histone

Query Match 35.7%; Score 5; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 55; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 9 GGSYR 13

|||||

Db 11 GGSYR 15

RESULT 12

B48411
herculin homolog - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: B48411
R;Neville, C.M.; Schmidt, M.; Schmidt, J.
Cell. Mol. Neurobiol. 12, 511-527, 1992
A;Title: Response of myogenic determination factors to cessation and resumption of elect
A;Reference number: A48411; MUID:93145310; PMID:1337017
A;Accession: B48411
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-64 <NEV>
A;Cross-references: GB:S53720; NID:g2644405; PID:g2644406
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:123817, NCBIP:123819)
C;Superfamily: human myogenin

Query Match 35.7%; Score 5; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 78; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 2 QRLPK 6

|||||

Db 50 QRLPK 54

RESULT 13

A48411
Myf5 homolog - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C;Accession: A48411
R;Neville, C.M.; Schmidt, M.; Schmidt, J.
Cell. Mol. Neurobiol. 12, 511-527, 1992
A;Title: Response of myogenic determination factors to cessation and resumption of elect
A;Reference number: A48411; MUID:93145310; PMID:1337017
A;Accession: A48411
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-64 <NEV>
A;Cross-references: GB:S53719; NID:g2644403; PID:g2644404
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:123813, NCBIP:123816)

C;Superfamily: human myogenin

Query Match 35.7%; Score 5; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 78; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 2 QRLPK 6

|||||

Db 50 QRLPK 54

RESULT 14

E98129
hypothetical protein AGR_L_17 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: E98129
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: E98129
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-65 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88559.1; PID:g15158266; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_17
A;Map position: linear chromosome

Query Match 35.7%; Score 5; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 79; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 10 GSYRC 14

|||||

Db 34 GSYRC 38

RESULT 15

S07516
gene 5.7 protein - phage T3
C;Species: phage T3
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 08-Oct-1999
C;Accession: S07516
R;Beck, P.J.; Gonzalez, S.; Ward, C.L.; Molineux, I.J.
J. Mol. Biol. 210, 687-701, 1989
A;Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.
A;Reference number: S07500; MUID:90133923; PMID:2614843
A;Accession: S07516
A;Molecule type: DNA
A;Residues: 1-69 <BEC>
A;Cross-references: EMBL:X17255; NID:g15682; PIDN:CAA35145.1; PID:g15707
C;Genetics:
A;Gene: 5.7

Query Match 35.7%; Score 5; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 83; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY 6 KRMGG 10

|||||

Db 63 KRMGG 67

Search completed: March 4, 2003, 17:02:33
Job time : 8.16949 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:56:44 ; Search time 4.27119 Seconds
(without alignments)
135.950 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 14

Sequence: 1 RQRLPRMGGSYRC 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	1162	1 TCNA_TRYCR	P23253 trypanosoma
2	6	42.9	146	1 RL15_BACSU	P19946 bacillus su
3	6	42.9	287	1 LEP4_XANCP	Q56763 xanthomonas
4	6	42.9	337	1 FXLI_MOUSE	Q64731 mus musculus
5	6	42.9	1104	1 SYV_YEAST	P07806 saccharomyc
6	5	35.7	54	1 RS14_AERPE	P58731 aeropyrum p
7	5	35.7	59	1 FLGC_BORHE	P70911 borrelia he
8	5	35.7	69	1 V57_EPT3	P20320 bacterioph
9	5	35.7	77	1 PUFQ_RHOSH	P16069 rhodobacter
10	5	35.7	94	1 PRIB_HAEIN	P44748 haemophilus
11	5	35.7	108	1 CPA3_CANPG	P81577 cancer pagu
12	5	35.7	122	1 RK14_CHLRE	P11094 chlamydomon
13	5	35.7	127	1 HOL2_HOLDI	Q25054 holotrichia
14	5	35.7	131	1 V14K_BSMV	P04869 barley stri
15	5	35.7	146	1 RL15_BACHD	P38373 bacillus ha
16	5	35.7	146	1 RL15_BACST	P04452 bacillus st
17	5	35.7	146	1 RL15_STAMP	O06445 staphylococ
18	5	35.7	147	1 RL15_THEMEA	Q9x1j0 thermotoga
19	5	35.7	150	1 RL15_RICPR	Q9zcs4 rickettsia
20	5	35.7	150	1 RL15_TETH	P74910 thermus the
21	5	35.7	156	1 RR7_NEPOL	Q9tkz6 nephroselm
22	5	35.7	156	1 V187_FOWPV	Q9j546 fowlpox vir
23	5	35.7	163	1 CIRP_XENLA	Q93235 xenopus lae
24	5	35.7	169	1 V55_EPT7	P03787 bacterioph
25	5	35.7	172	1 CIRP_HUMAN	Q14011 homo sapien
26	5	35.7	172	1 CIRP_MOUSE	Q61413 mus musculus
27	5	35.7	176	1 YW1_CAEEL	Q11088 caenorhabdi
28	5	35.7	180	1 RL15_LEPIN	Q9xd17 leptospira
29	5	35.7	184	1 Y807_METJA	Q58217 methanococ
30	5	35.7	191	1 RL15_PTRAJ	Q8zwd8 pyrobaculum
31	5	35.7	206	1 PAAD_SYNY3	P72743 synchocyst
32	5	35.7	210	1 YQEH_ECOLI	Q46941 escherichia
33	5	35.7	215	1 RL15_SULTO	Q975g1 sulfolobus

ALIGNMENTS

RESULT 1

ID	TCNA_TRYCR	STANDARD;	PRT;	1162 AA.
AC	P23253;			
DT	01-NOV-1991 (Rel. 20, Created)			
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen).			
GN	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]_TaxID=5693;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Silvio X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,			
RA	Prioli R.P.;			
RT	"The Trypanosoma cruzi neuraminidase contains sequences similar to			
RT	bacterial neuraminidases, YWTD repeats of the low density lipoprotein			
RT	receptor, and type III modules of fibronectin."			
RL	J. Exp. Med. 174:179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION			
RX	MEDLINE=91376547; PubMed=1896773;			
RA	Prioli R.P., Mejia J.S., Aji T., Aikawa M., Pereira M.E.A.;			
RT	"Trypanosoma cruzi: Localization of neuraminidase on the surface of			
RT	trypomastigotes."			
RL	Trop. Med. Parasitol. 42:146-150(1991).			
CC	-!- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN			
CC	PARASITE INVASION OF CELLS.			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in			
CC	oligosaccharides, glycoproteins, glycolipids, colominic acid and			
CC	synthetic substrates.			
CC	-!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR			
CC	(POSSIBLE).			
CC	-!- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM			
CC	IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.			
CC	-!- MISCELLANEOUS: THE VARIABLE LENGTHS OF THE LONG TANDEM REPEAT			
CC	DOMAIN COULD ACCOUNT IN PART FOR THE POLYMORPHISM OF THE TCNA			
CC	PROTEIN.			
CC	-!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.			
CC	-!- SIMILARITY: CONTAINS 3 BNR REPEATS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M61732; AAA30255.1; -.			
DR	PIR; JH0557; JH0557.			
DR	InterPro; IPR002860; GH_BNR.			

34	5	35.7	216	1	HAN1_MOUSE	Q64279 m heart- an
35	5	35.7	216	1	R15E_SULSO	Q9uxd0 sulfolobus
36	5	35.7	224	1	MYOG_HUMAN	P15173 homo sapien
37	5	35.7	224	1	MYOG_MOUSE	P12979 mus musculu
38	5	35.7	224	1	MYOG_PIG	P49812 sus scrofa
39	5	35.7	225	1	R15E_AERPE	Q9yb28 aeropyrum p
40	5	35.7	226	1	BIOD_XYLEA	Q9pal9 xyella fas
41	5	35.7	226	1	MYOG_COTUA	P34060 coturnix co
42	5	35.7	227	1	MYOG_CHICK	P17920 gallus gall
43	5	35.7	228	1	TIG1_HUMAN	P49788 homo sapien
44	5	35.7	236	1	RR2_ARATH	P56797 arabidopsis
45	5	35.7	236	1	RR2_SPIOL	P08242 spinacia ol

DR Pfam; PF02012; BNR; 2.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 588 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 1162;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14
|||||
DB 379 RQRLPKRMGGSYRC 392

RESULT 2
RL15 BACSU STANDARD; PRT; 146 AA.
AC P19946; 1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L15.
GN RPL0.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90292990; PubMed=2113521;
RA Nakamura K., Nakamura A., Takamatsu H., Yoshikawa H., Yamane K.;
RT "Cloning and characterization of a Bacillus subtilis gene homologous
RT to E. coli secY.";
RL J. Biochem. 107:603-607(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90221911; PubMed=2139212;
RA Yoshikawa H., Doi R.H.;
RT "Sequence of the Bacillus subtilis spectinomycin resistance gene
RT region.";
RL Nucleic Acids Res. 18:1647-1647(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=96186897; PubMed=8635744;
RA Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT "Genetic and transcriptional organization of the Bacillus subtilis
RT spc-alpha region.";
RL Gene 169:17-23(1996).
RN [4]
RP SEQUENCE OF 94-146 FROM N.A.
RC STRAIN=168;
RX MEDLINE=90251170; PubMed=2110998;
RA Suh J.-W., Boylan S.A., Thomas S.M., Dolan K.M., Oliver D.B.,
RA Price C.W.;
RT "Isolation of a secY homologue from Bacillus subtilis: evidence for a
RT common protein export pathway in eubacteria.";
RL Mol. Microbiol. 4:305-314(1990).
CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -!- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00619; BAA00494.1; -;
DR EMBL; X51329; CAA35711.1; -;
DR EMBL; L47971; AAB06818.1; -;
DR EMBL; M31102; AAB59117.1; -;
DR EMBL; Z99104; CAB11911.1; -;
DR PIR; S12682; R5BSL5.
DR Subtilist; BG10444; rplO.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR Pfam; PF01305; Ribosomal_L15; 1.
DR TIGRfams; TIGR01071; rplO_bact; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 146 AA; 15383 MW; 5DB07A902B266C11 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRLPKR 7
|||||
DB 59 QRLPKR 64

RESULT 3
LEP4 XANCP STANDARD; PRT; 287 AA.
AC Q56763;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Type 4 prepilin-like proteins leader peptide processing enzyme
DE [Includes: Leader peptidase (EC 3.4.99.-) (Prepilin peptidase); N-
DE methyltransferase (EC 2.1.1.-)].
GN XPSO OR PILD OR XCC3101.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xc1701;
RX MEDLINE=96414476; PubMed=8817497;
RA Hu N.-T.T., Lee P.F., Chen C.;
RT "The type IV pre-pilin leader peptidase of Xanthomonas campestris pv.
RT campestris is functional without conserved cysteine residues.";
RL Mol. Microbiol. 18:769-777(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.P.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chamergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fomighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).

CC -1- FUNCTION: CLEAVES TYPE-4 FIMBRIAL LEADER SEQUENCE AND METHYLATES
 CC THE N-TERMINAL (GENERALLY PHE) RESIDUE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A24.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U12432; AAC43571.1; -.
 DR EMBL; AE012425; AAM42372.1; -.
 DR MEROPS; A24.001; -.
 DR InterPro; IPR000045; Peptidase C20.
 DR Pfam; PF01478; Peptidase C20; 1.
 DR PRINTS; PR00864; PREPILNPFASE.
 KW Multifunctional enzyme; Hydrolase; Protease; Transferase;
 KW Methyltransferase; Transmembrane; Inner membrane.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 101 121 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 253 273 POTENTIAL.
 SQ SEQUENCE 287 AA; 31843 MW; A58DD1A514F9ECC2 CRC64;
 Query Match 42.9%; Score 6; DB 1; Length 287;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 3 RLPRKM 8
 Db 32 RLPRKM 37
 RESULT 4
 FXL1 MOUSE
 ID _FXL1 MOUSE STANDARD; PRT; 337 AA.
 AC Q84731;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Forkhead box protein 1l (Forkhead-related protein FKHL1l)
 DE (Transcription factor FKX-6).
 GN FOXL1 OR FKHL1 OR FKX6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=96232289; PubMed=8674414;
 RA Kaestner K.H., Bleckmann S.C., Monaghan A.P., Schlondorff J.,
 RA Mincheva A., Lichter P., Schuetz G.;
 RT "Clustered arrangement of winged helix genes fkh-6 and MFH-1:
 RT possible implications for mesoderm development.";
 RL Development 122:1751-1758 (1996).
 RN [2]
 RP SEQUENCE OF 40-150 FROM N.A.
 RC STRAIN=129;
 RX MEDLINE=93361500; PubMed=7689224;
 RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H., Monaghan A.P.,
 RA Schuetz G.;
 RT "Six members of the mouse forkhead gene family are developmentally
 RT regulated";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631 (1993).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X92498; CAA63243.1; -.
 DR EMBL; X71944; CAA50746.1; -.
 DR HSPSP; Q63245; 2HPH.
 DR TRANSFAC; T02417; -.
 DR MGD; MGI:1347469; Foxl1.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR ProDom; PD000425; TF_Fork_head; 1.
 DR SMART; SMO0339; FH; 1.
 DR PROSITE; PS00657; FORK HEAD 1; 1.
 DR PROSITE; PS00658; FORK HEAD 2; 1.
 DR PROSITE; PS50039; FORK HEAD 3; 1.
 KW DNA-binding; Nuclear protein; Transcription regulation.
 FT DNA_BIND 48 139
 SQ SEQUENCE 337 AA; 36541 MW; 82FBD018B81DD6BD CRC64;
 Query Match 42.9%; Score 6; DB 1; Length 337;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 QRLPKR 7
 Db 244 QRLPKR 249
 RESULT 5
 SYV_YEAST
 ID _SYV_YEAST STANDARD; PRT; 1104 AA.
 AC P07806;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Valyl-tRNA synthetase, mitochondrial precursor (EC 6.1.1.9)
 DE (Valine-tRNA ligase) (ValRS).
 GN VAS1 OR YGR094W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87222321; PubMed=3294828;
 RA Jordana X., Chatton B., Paz-Weisshaar M., Buhler J.-M., Cramer F.,
 RA Ebel J.-P., Fasiolo F.;
 RT "Structure of the yeast valyl-tRNA synthetase gene (VAS1) and the
 RT homology of its translated amino acid sequence with Escherichia coli
 RT isoleucyl-tRNA synthetase.";
 RL J. Biol. Chem. 262:7189-7194 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Wedler H., Scharfe M., Wedler E., Wambutt R.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hernandez K., Weber N., Wipfli P., Schmidheini T.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-78 FROM N.A.
 RX MEDLINE=88087140; PubMed=3275649;
 RA Chatton B., Walter P., Ebel J.-P., Lacroute F., Fasiolo F.;
 RT "The yeast VAS1 gene encodes both mitochondrial and cytoplasmic
 RT valyl-tRNA synthetases";
 RL J. Biol. Chem. 263:52-57 (1988).


```

SQ SEQUENCE 59 AA; 6571 MW; B2C27B6B4C238D04 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYR 13
Db 39 GGSYR 43

RESULT 8
V57_BPT3
ID V57_BPT3 STANDARD; PRT; 69 AA.
AC P20320;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Gene 5.7 protein.
GN 5.7.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Luria;
RX MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL J. Mol. Biol. 210:687-701(1989).
CC -!- FUNCTION: ALLOWS GROWTH ON LAMBDA LYSOGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; X17255; CAA35145.1; -.
DR PIR; S07516; S07516.
SQ SEQUENCE 69 AA; 7381 MW; EDDP2572C7086FC9 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 69;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KRMGG 10
Db 63 KRMGG 67

RESULT 9
PUFQ_RHOSH
ID PUFQ_RHOSH STANDARD; PRT; 77 AA.
AC P16069;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein pufQ.
GN PUFQ.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC13;
RX MEDLINE=92140030; PubMed=1779756;
RA Hunter C.N., McGlynn P., Ashby M.K., Burgess J.G., Olsen J.D.;
RT "DNA sequencing and complementation/deletion analysis of the bchA-puf
operon region of Rhodobacter sphaeroides: in vivo mapping of the
oxygen-regulated puf promoter.";
Mol. Microbiol. 5:2649-2661(1991).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=94245622; PubMed=8188596;
RA Gong L., Lee J.K., Kaplan S.;
RT "The Q gene of Rhodobacter sphaeroides: its role in puf operon
expression and spectral complex assembly.";
J. Bacteriol. 176:2946-2961(1994).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
sphaeroides 2.4.1.";
Nucleic Acids Res. 28:862-867(2000).
[4]
RP SEQUENCE OF 1-7 FROM N.A.
RC MEDLINE=93173096; PubMed=8437569;
RA McGlynn P., Hunter C.N.;
RT "Genetic analysis of the bchC and bchA genes of Rhodobacter
sphaeroides.";
Mol. Gen. Genet. 236:227-234(1993).
[5]
RP SEQUENCE OF 44-77 FROM N.A.
RX MEDLINE=87109067; PubMed=3027044;
RA Kiley P.J., Donohue T.J., Haveika W.A., Kaplan S.;
RT "DNA sequence and in vitro expression of the B875 light-harvesting
polypeptides of Rhodobacter sphaeroides.";
J. Bacteriol. 169:742-750(1987).
CC -!- FUNCTION: REQUIRED FOR BACTERIOCHLOROPHYLL BIOSYNTHESIS. DIRECTLY
CC INVOLVED IN THE ASSEMBLY OF BOTH THE B875 AND B800-850 PIGMENT-
CC PROTEIN COMPLEXES.
CC -!- SIMILARITY: BELONGS TO THE PUFQ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AJ010302; CAB38750.1; -.
DR ENBL; L25894; AAD15241.1; -.
DR ENBL; AF195122; AAF24300.1; -.
DR ENBL; M15105; -; NOT_ANNOTATED_CDS.
DR PIR; S18581; S18581.
KW Photosynthesis; Chlorophyll biosynthesis.
FT CONFLICT 36 36 A -> G (IN REF. 1).
SQ SEQUENCE 77 AA; 8656 MW; 61127EF670B39E86 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKR 7
Db 53 RLPKR 57

RESULT 10
PRIB_HAEIN
ID PRIB_HAEIN STANDARD; PRT; 94 AA.
AC P44748;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Primosomal replication protein N.
GN PRIB OR HI0546.

```

```
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlavagne A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Karlavagne A.R., Fleischmann R.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX
CC SSDNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS
CC PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE
CC SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; U32736; AAC22204.1; -.
DR TIGR; H10546; -.
KW Primosome; Complete proteome.
SQ SEQUENCE 94 AA; 10529 MW; 6D639677E2DCC582 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKR 7
DB 6 RLPKR 10

RESULT 11
CPA3 CANPG STANDARD; PRT; 108 AA.
AC P81577;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cuticle protein AM1199 (CPAM1199).
OS Cancer pagurus (Rock crab).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Brachyura; Eubrachyura; Cancroidea; Cancridae; Cancer.
OX NCBI_TaxID=6755;
RN [1]
RP SEQUENCE.
RC TISSUR=Cuticle;
RX MEDLINE=99354472; PubMed=10425740;
RA Andersen S.O.;
RT "Exoskeletal proteins from the crab, Cancer pagurus.";
RL Comp. Biochem. Physiol. 123A:203-211(1999).
CC -1- TISSUE SPECIFICITY: ARTHRODIAL MEMBRANE.
```

```
CC -1- MASS SPECTROMETRY: MW=12190.5; METHOD=MALDI.
CC -1- SIMILARITY: CONTAINS 1 CUTICLE CONSENSUS DOMAIN.
DR InterPro; IPR000618; Insect cuticle.
DR Pfam; PF00379; Chitin bind_4; 1.
DR PRINTS; PR00947; CUTICLE.
DR PROSITE; PS00233; CUTICLE; 1.
KW Structural protein; Cuticle; Glycoprotein.
FT MOD_RES 1 PYROLIDONE CARBOXYLIC ACID.
FT CARBOHYD 89 O-LINKED (HEXNAC).
SQ SEQUENCE 108 AA; 12011 MW; 768F71C85F1BBCFF CRC64;

Query Match 35.7%; Score 5; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYR 13
DB 56 GGSYR 60

RESULT 12
RK14 CHLRE STANDARD; PRT; 122 AA.
ID RK14_CHLRE
AC P11094;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L14.
GN RPL14.
OS Chlamydomonas reinhardtii.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=137C / CC-125;
RX MEDLINE=89263799; PubMed=2726491;
RA Lou J.K., Cruz F.D., Wu M.;
RT "Nucleotide sequence of the chloroplast ribosomal protein gene L14 in
RL Chlamydomonas reinhardtii."
CC -1- SIMILARITY: BELONGS TO THE L14P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; X14062; CAA32226.1; -.
DR PIR; S04093; RSKM14.
DR HSP; P04450; LWHI.
DR InterPro; IPR000218; Ribosomal L14.
DR Pfam; PF00238; Ribosomal L14; 1.
DR ProDom; PD001093; Ribosomal L14; 1.
DR TIGRfams; TIGR01067; rplN_bact; 1.
DR PROSITE; PS00049; RIBOSOMAL_L14; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 122 AA; 13446 MW; 3C66C3C05D15FE76 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGSYR 13
DB 26 GGSYR 30

RESULT 13
```



```
HOL2_HOLDI STANDARD; PRT; 127 AA.
AC Q25054;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Holotricin 2 precursor.
OS Holotrichia diomphalia.
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Scarabaeiformia; Scarabaeidae; Melolonthinae; Holotrichia.
OX NCBI_TaxID=33394;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 56-75 AND 82-126.
RC TISSUE=Larval hemolymph;
RA MEDLINE=94245669; PubMed=8188641;
RA Lee S.Y., Moon H.J., Kurata S., Kurama T., Natori S., Lee B.L.;
RT "Purification and molecular cloning of cDNA for an inducible
RT antibacterial protein of larvae of a coleopteran insect, Holotrichia
RT diomphalia."
RL J. Biochem. 115:82-86(1994).
CC -I- FUNCTION: ANTIBACTERIAL ACTIVITY AGAINST GRAM-NEGATIVE BACTERIA
CC BUT NOT AGAINST GRAM-POSITIVE BACTERIA.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- TISSUE SPECIFICITY: HEMOLYMPH.
CC -I- SIMILARITY: BELONGS TO THE COLEOPTERICIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13745; BAA02890.1; --
KW Signal; Insect immunity; Antibiotic; Hemolymph.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 55
FT CHAIN 56 127 HOLOTRICIN 2.
SQ SEQUENCE 127 AA; 14075 MW; 3478B6DD308AC35B CRC64;
Query Match 35.7%; Score 5; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GGSYR 13
DB 122 GGSYR 126
RESULT 14
V14K_BSMV STANDARD; PRT; 131 AA.
AC P04869;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE 14 kDa protein (Beta-C protein).
OS Barley stripe mosaic virus (BSMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Hordeivirus.
OX NCBI_TaxID=12327;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC PV43;
RX MEDLINE=86232627; PubMed=3754962;
RA Gustafson G., Armour S.L.;
RT "The complete nucleotide sequence of RNA beta from the type strain of
RT barley stripe mosaic virus."
RL Nucleic Acids Res. 14:3895-3909(1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03854; CAA27486.1; --
DR PIR; A04192; MMBV4B.
DR InterPro; IPR001896; Plant_vir_prot.
DR Pfam; PF01307; Plant_vir_prot; 1.
DR ProDom; PD001561; Plant_vir_prot; 1.
SQ SEQUENCE 131 AA; 14119 MW; 7106177BF6230B43 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 131;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 GGSYR 13
DB 49 GGSYR 53
RESULT 15
RL15_BACHD STANDARD; PRT; 146 AA.
AC P38373; Q9WMJ2; Q9JPM8;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE S0S ribosomal protein L15.
OS Bacillus halodurans.
GN RPL0 OR BH0153.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99209008; PubMed=10192928;
RA Takami H., Takaki Y., Nakasone K., Hiramata C., Inoue A., Horikoshi K.;
RT "Sequence analysis of a 32-kb region including the major ribosomal
RT protein gene clusters from alkaliphilic Bacillus sp. strain C-125."
RL Biosci. Biotechnol. Biochem. 63:452-455(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
RN [3]
RP SEQUENCE OF 81-146 FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=92381482; PubMed=1512566;
RA Kang S.K., Kudo T., Horikoshi K.;
RT "Molecular cloning and characterization of an alkaliphilic Bacillus
RT sp. C125 gene homologous to Bacillus subtilis secY."
RL J. Gen. Microbiol. 138:1365-1370(1992).
CC -I- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA.
CC -I- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017508; BAA75290.1; --
DR EMBL; AF001507; BAB03872.1; --
```

DR EMBL; D10360; BAA01190.1; -.
DR PIR; A44859; A44859.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR Pfam; PF01305; Ribosomal_L15; 1.
DR TIGRFAMS; TIGR01071; rplO_bact; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 146 AA; 15534 MW; F67DA91D15AC9048 CRC64;

Query Match 35.78; Score 5; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPKR 7
Db 60 RLPKR 64
|||||
|||||

Search completed: March 4, 2003, 17:00:32
Job time : 5.27119 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:57:01 ; Search time 11.6271 Seconds
(without alignments)
248.098 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 14

Sequence: 1 RORLPKRMGGSYRC 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6	42.9	138	17	O26189 methanobact
2	6	42.9	166	4	Q9NXL1 homo sapien
3	6	42.9	207	15	Q72475 human immun
4	6	42.9	232	15	Q9J810 human immun
5	6	42.9	323	2	Q8RN81 aeromonas h
6	6	42.9	334	5	Q22298 caenorhabdi
7	6	42.9	338	2	Q9F580 escherichia
8	6	42.9	447	5	Q27908 pinctada fu
9	6	42.9	676	5	Q9VED0 drosophila
10	6	42.9	717	10	O49671 arabisidopsis
11	6	42.9	755	10	Q9FGM2 arabisidopsis
12	6	42.9	992	5	Q9VM31 drosophila
13	6	42.9	1101	16	Q9I755 pseudomonas
14	5	35.7	46	4	Q9P027 homo sapien
15	5	35.7	50	4	Q8TDB0 homo sapien
16	5	35.7	50	10	Q8S414 zea mays (m

ALIGNMENTS

RESULT 1

O26189 ID O26189 PRELIMINARY; PRT; 138 AA.
AC O26189;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein MTH86.
GN MTH86.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463; DeLoughery C., Lee H.-M., Dubois J., Smith D.R., Doucette-Stamm L.A., Blakely D., Cook R., Gilbert K., Aldredge T., Bashirzadeh R., Keagle P., Lumm W., Pothier B., Qiu D., Harrison D., Hoang L., Wang Y., Wierzbowski J., Gibson R., Spadafora R., Vicare R., Bush D., Safer H., Patwell D., Prabhakar S., Jiwan N., Caruso A., Shimer G., Goyal A., Pietrowski S., Church G.M., McDougall S., Mao J.-I., Rice P., Noelling J., Reeve J.N.; Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.; "Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics."; J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000799; AAB84592.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 138 AA; 15306 MW; 29EC2708F1496A5A CRC64;

Query Match 42.9%; Score 6; DB 17; Length 138;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KRMGGS 11

Db 8 KRMGGS 13

RESULT 2

Q9NXB1 ID Q9NXB1 PRELIMINARY; PRT; 166 AA.
AC Q9NXB1
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA FLJ20344 fis, clone HEP13661.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Oka T., Suzuki Y., Obayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegai T., Sugano S.,
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000351; BAA91104.1; -
DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS0805; KRAB; 1.
SQ SEQUENCE 166 AA; 19473 MW; 529505DDAE26870D CRC64;

Query Match 42.9%; Score 6; DB 4; Length 166;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RORLPK 6
|
|
|
|
|
|
Db 135 RORLPK 140

RESULT 3

Q72475 ID Q72475 PRELIMINARY; PRT; 207 AA.
AC Q72475
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Negative factor (F-protein) (27 kDa protein).
GN NEF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9601370; PubMed=7474087;
RA Michael N.L., Chang G., d'Arcy L.A., Tseng C.J., Birx D.L.,
RA Sheppard H.W.;
RT "Functional characterization of human immunodeficiency virus type 1
nef genes in patients with divergent rates of disease progression.";
RL J. Virol. 69:6758-6769(1995).
CC -!- SIMILARITY: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
CC ACTIVITIES, IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN.
DR EMBL; U26136; AAA87539.1; -
DR HSP; P03406; 1EFN.
DR InterPro; IPR001558; HIV Nef.
DR Pfam; PF00469; F-protein; 1.
DR ProDom; PD000031; HIV Nef; 1.
KW AIDS; GTP-binding; Myristate.
SQ SEQUENCE 207 AA; 23630 MW; 0264A0C50934CAC1 CRC64;

Query Match 42.9%; Score 6; DB 15; Length 207;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PKRMGG 10
|
|
|
|
|
|
Db 6 PKRMGG 11

RESULT 4

Q9J8L0 ID Q9J8L0 PRELIMINARY; PRT; 232 AA.
AC Q9J8L0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pol protein (Fragment).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1.95C19;
RX MEDLINE=20120932; PubMed=10653914;
RA Briones C., Mas A., Gomez-Mariano G., Altisent C., Menendez-Arias L.,
RA Soriano V., Domingo E.;
RT "Dynamics of dominance of a dipeptide insertion in reverse
transcriptase of HIV-1 from patients subjected to prolonged therapy.";
RL Virus Res. 66:13-26(2000).
DR EMBL; AF168273; AAF44226.1; -
DR HSP; P16088; 1FIV.
DR InterPro; IPR001995; Aspprotease rtrv.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00077; rvp; 1.
DR Pfam; PF00078; rvt; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 232
SQ SEQUENCE 232 AA; 25833 MW; 368C461F4242AAAA CRC64;

Query Match 42.9%; Score 6; DB 15; Length 232;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PKRMGG 10
|
|
|
|
|
|
Db 45 PKRMGG 50

RESULT 5

Q8RN81 ID Q8RN81 PRELIMINARY; PRT; 323 AA.
AC Q8RN81
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aerolysin (Fragment).
GN AERA.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AHK6;
RA Ahmad A., Jangi S., Usup G.;
RT "High prevalence of beta-hemolytic strains among clinical and
environmental isolates of Aeromonas hydrophila from Malaysia.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF485768; AAL91586.1; -
FT NON_TER 1
FT NON_TER 323
SQ SEQUENCE 323 AA; 35890 MW; CA0EAF020E223921 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 323;
Best Local Similarity 100.0%; Pred. No. 47;

Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
---------	----	--------------	----	------------	----	--------	----	------	----

Qy 9 GGSYRC 14
|||
Db 23 GGSYRC 28

RESULT 6

Q22298	PRELIMINARY;	PRT;	334 AA.
ID	Q22298		
AC	Q22298;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	T07C5.2 protein.		
GN	T07C5.2.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Pelodexinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	McMurray A.A.;		
RL	Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.		
RN	[2]		

RESULT 7

Q9F580	PRELIMINARY;	PRT;	338 AA.
ID	Q9F580		
AC	Q9F580;		
DT	01-MAR-2001	(TrEMBLrel. 16, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last annotation update)	
DE	Possible replication initiation protein.		
GN	REPA.		
OS	Escherichia coli.		
OG	Plasmid R721.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K-12; TRANSPOSON=TN7;		

RESULT 9

Q9VED0
ID Q9VED0
AC Q9VED0;
PRELIMINARY;
PRT; 676 AA.

RA Sampei G., Motomura K., Masuda S., Yamauchi T., Ando K., Oishi T.,
RC STRAIN-K-12; TRANSPOSON-TN7;
RD MEDLINE=93015772; PubMed=1400257;
RE Kim S., Komano T.;
RF "Nucleotide sequence of the R721 shufflon.";
RG J. Bacteriol. 174:7053-7058(1992).
RH EMBL; AF002527; BAB12585.1; -.
RI Plasmid.
RJ
RK
RL
RM
RN [2]
RO SEQUENCE FROM N.A.
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
SZ

RESULT 8

```

QZ27908      PRELIMINARY;      PRT;      447 AA.
AC Q27908
AD Q27908;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE De Nacrein
DE OS
OS Pinctada fucata.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Periomorpha; Perioidea;
OC Pterioidea; Pteriidae; Pinctada.
OX NCBI_TaxId=50426;
[1]
RN SEQUENCE FROM N.A.
RP RP
RC TISSUE=MANTLE;
RA Miyamoto H.;
RL "A carbonic anhydrase from the nacreous layer in oyster pearls.";
RT Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1996).
RN [2]
RN SEQUENCE FROM N.A.
RP RP
RC TISSUE=MANTLE;
RC MEDLINE=96382522; PubMed=8790386;
RX Miyamoto H., Miyashita T., Okushima M., Nakano S., Morita T.,
RX Matsushiro A.;
RT "A carbonic anhydrase from the nacreous layer in oyster pearls.";
RT Proc. Natl. Acad. Sci. U.S.A. 93:9657-9660(1996).
RL EMBL; D83523; BAA11940.1; -.
DR HSSP; P22748; 12NC.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001148; Euk_Coahnd.
DR Pfam; PF00194; carb_anhydrase; 3.
DR Pfam; PF01391; Collagen; 1.
DR ProDom; PD000865; Euk_Coahnd; 2.
SO SEQUENCE      447 AA; 50114 MW; 334191045E931AB8C CRC64;

```

RESULT 9

Q9VED0
ID Q9VED0
AC Q9VED0;
PRELIMINARY;
PRT; 676 AA.

01-MAY-2000 (TrEMBLrel. 13, Created)
 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 CG7218 protein (GH03607p).
 CG7218.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=2019606; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003721; AAF55495.2; --
 DR EMBL; AY058285; AAL13544.7; --
 DR Flybase; FBgn0038569; CG7218.
 SQ SEQUENCE 676 AA; 75565 MW; 26279B9B2711971D6 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 676;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ORLPKR 7
 Db 639 ORLPKR 644
 |||||

RESULT 10

049671 PRELIMINARY; PRT; 717 AA.
 ID 049671
 AC 049671;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Replication A protein-like.
 GN T18B16.100 OR A4GL9130.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoerge W., Bancroft I.,
 RA Mewes H.W., Mayer K., Schueller C.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021687; CAA16702.1; --
 DR EMBL; AL161550; CAB78915.1; --
 DR HSSP; P27694; 1JMC
 DR InterPro; IPR004591; Rpal.
 DR InterPro; IPR004365; tRNA anti.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01336; tRNA anti; 2.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00343; ZNF_C2HC; 1.
 DR TIGRFAMs; TIGR00617; rpal; 1.
 SQ SEQUENCE 717 AA; 78942 MW; 4E005B0184E0A816 CRC64;

Query Match 42.9%; Score 6; DB 10; Length 717;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 MGGSYR 13
 Db 694 MGGSYR 699
 |||||

RESULT 11

049671 PRELIMINARY; PRT; 755 AA.
 ID 049671
 AC 049671;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE DnaJ protein-like.
 GN DnaJ protein-like.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB025622; BAB08418.1; --

```
DR HSP; P25685; 1HDJ.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR PRINTS; PR00625; DnaJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; UNKNOWN_1.
DR PROSITE; PS0076; DnaJ_2; 1.
SQ SEQUENCE 755 AA; 85237 MW; 47CS9BDBDF62560FE CRC64;

Query Match 42.9%; Score 6; DB 10; Length 755;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMGGSY 12
   |||||
Db 294 RMGGSY 299

RESULT 12
Q9VM91 PRELIMINARY; PRT; 992 AA.
ID Q9VM91
AC Q9VM91
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG11323 protein.
GN CG11323.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellista A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"the genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
DR EMBL; AR003614; AAF52432.1; -.
DR FlyBase; FBgn0031854; CG11323.
DR InterPro; IPR004344; Tub_tyr_lygase.
DR Pfam; PF03133; TTL; 1.
SQ SEQUENCE 992 AA; 112255 MW; 0FC34674C15BDAAD CRC64;

Query Match 42.9%; Score 6; DB 5; Length 992;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QRLPKR 7
   |||||
Db 186 QRLPKR 191

RESULT 13
Q9I755 PRELIMINARY; PRT; 1101 AA.
ID Q9I755
AC Q9I755
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein PA0077.
GN PA0077.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AR004446; AAG03467.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1101 AA; 121458 MW; 216CCAD5E6F768CE CRC64;

Query Match 42.9%; Score 6; DB 16; Length 1101;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RORLPK 6
   |||||
Db 583 RORLPK 588

RESULT 14
Q9P027 PRELIMINARY; PRT; 46 AA.
ID Q9P027
AC Q9P027
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Latent transforming growth factor beta binding protein 3
DE (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Saharinen J., Penttinen C., Keski-Oja J.;
RT "Cloning of human LTBP-3.";
```

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF135961; AAF62353.2; -;
 DR HSP; P35555; IEMN.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001438; EGF_II.
 DR Pfam; PF00008; EGF; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR SMART; SM00181; EGF; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 46
 SQ SEQUENCE 46 AA; 4971 MW; 211C501D8B844A8C CRC64;

Query Match 35.7%; Score 5; DB 4; Length 46;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GSYRC 14
 Db 20 GSYRC 24

RESULT 15

Q8TDB0 PRELIMINARY; PRT; 50 AA.
 AC Q8TDB0
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Apolipoprotein A-1 A175P variant (Fragment).
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Booth D., Bybee A., Pepys M., Hawkins P.;
 RT "Hereditary amyloidosis."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF485255; AAL92035.1; -;
 KW Lipoprotein.
 FT NON_TER 1
 FT NON_TER 50
 SQ SEQUENCE 50 AA; 5607 MW; A65219AE9E965732 CRC64;

Query Match 35.7%; Score 5; DB 4; Length 50;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RORLP 5
 Db 17 RORLP 21

Search completed: March 4, 2003, 17:01:43
 Job time : 26.6271 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:58:06 ; Search time 5.45763 Seconds
(without alignments)
75.476 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 14

Sequence: 1 RORLPKMGSGYRC 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/prodata/1/iaa/PGTUS_COMB.pep.*

6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	42.9	430	2	US-08-945-848-8
2	5	35.7	15	3	Sequence 8, Appli
3	5	35.7	16	2	Sequence 8, Appli
4	5	35.7	16	2	Sequence 9, Appli
5	5	35.7	16	4	Sequence 9, Appli
6	5	35.7	41	6	Patent No. 5177197
7	5	35.7	45	1	Sequence 25, Appli
8	5	35.7	45	2	Sequence 24, Appli
9	5	35.7	66	2	Sequence 224, App
10	5	35.7	66	3	Sequence 224, App
11	5	35.7	66	4	Sequence 224, App
12	5	35.7	77	1	Sequence 1, Appli
13	5	35.7	77	1	Sequence 14, Appli
14	5	35.7	77	1	Sequence 1, Appli
15	5	35.7	77	2	Sequence 1, Appli
16	5	35.7	77	3	Sequence 14, Appli
17	5	35.7	77	3	Sequence 1, Appli
18	5	35.7	96	1	Sequence 5, Appli
19	5	35.7	146	4	Sequence 494, App
20	5	35.7	150	4	Sequence 3280, App
21	5	35.7	203	4	Sequence 383, App
22	5	35.7	224	4	Sequence 10, Appli
23	5	35.7	227	6	Patent No. 5428012
24	5	35.7	227	6	Patent No. 5428012
25	5	35.7	227	6	Patent No. 5428012
26	5	35.7	228	1	Sequence 5, Appli
27	5	35.7	228	6	Patent No. 5428012

28	5	35.7	228	6	5451506-2	Patent No. 5451506
29	5	35.7	252	1	US-08-792-019B-13	Sequence 13, Appli
30	5	35.7	252	3	US-08-988-819-13	Sequence 13, Appli
31	5	35.7	252	4	US-09-016-534-13	Sequence 13, Appli
32	5	35.7	252	4	US-08-097-869-3	Sequence 3, Appli
33	5	35.7	273	2	US-08-403-853-18	Sequence 18, Appli
34	5	35.7	282	4	US-09-455-960-9	Sequence 9, Appli
35	5	35.7	410	6	5177197-1	Patent No. 5177197
36	5	35.7	437	2	US-08-935-450-8	Sequence 2, Appli
37	5	35.7	463	4	US-09-163-444-2	Sequence 2, Appli
38	5	35.7	468	1	US-08-220-151-16	Sequence 16, Appli
39	5	35.7	468	1	US-08-413-118-16	Sequence 16, Appli
40	5	35.7	468	3	US-08-473-446-16	Sequence 16, Appli
41	5	35.7	485	1	US-08-344-536-4	Sequence 4, Appli
42	5	35.7	485	3	US-08-920-562-4	Sequence 4, Appli
43	5	35.7	511	4	US-08-991-677-4	Sequence 4, Appli
44	5	35.7	560	2	US-08-808-931-18	Sequence 18, Appli
45	5	35.7	560	3	US-08-808-323-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-08-945-848-8

; Sequence 8, Application US/08945848

; Patent No. 5968772

; GENERAL INFORMATION:

; APPLICANT: MATSUSHIRO, Aizo

; TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR

; TITLE OF INVENTION: THE SAME

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 233 South Wacker Drive/6300 Sears Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/945,848

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Cawley, Jr, Thomas A.

; REGISTRATION NUMBER: 40,944

; REFERENCE/DOCKET NUMBER: 19036/34324

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (312) 474-6300

; TELEFAX: (312) 474-0448

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 430 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-945-848-8

Query Match 42.9%; Score 6; DB 2; Length 430;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RORLPK 6

Db 77 RORLPK 82

```

--
RESULT 2
US-09-067-284-8
; Sequence 8, Application US/09067284
; Patent No. 6071691
; GENERAL INFORMATION:
; APPLICANT: Hoesktra, Merl F.
; APPLICANT: Thayer, Mathew J.
; TITLE OF INVENTION: Materials and Methods for Modulating
; TITLE OF INVENTION: Differentiation
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive; 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067.284
; FILING DATE: April 27, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/34011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-067-284-8

Query Match 35.7%; Score 5; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRLPK 6
Db 11 QRLPK 15

RESULT 3
US-08-396-452-9
; Sequence 9, Application US/08396452
; Patent No. 5820863
; GENERAL INFORMATION:
; APPLICANT: Dunbar, Bonita S.
; TITLE OF INVENTION: Method of Preparation and Use for Zona
; TITLE OF INVENTION: Pellucida Antigens and Antibodies for Sterilization and
; TITLE OF INVENTION: Contraception
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1177 West Loop South, 10th Floor
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77027
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,119
; FILING DATE: 09-Oct-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0165
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..16
; OTHER INFORMATION: /note= "N-Terminal Amino Acid
; Sequence of Pig ZP Glycoprotein II(E)"

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,452
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/055,831
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0909
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-396-452-9

Query Match 35.7%; Score 5; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 RMGGS 11
Db 12 RMGGS 16

RESULT 4
US-09-169-119-9
; Sequence 9, Application US/09169119
; Patent No. 6264953
; GENERAL INFORMATION:
; APPLICANT: Dunbar, Bonita S.
; TITLE OF INVENTION: Method of Preparation and Use for Zona
; TITLE OF INVENTION: Pellucida Antigens and Antibodies for Sterilization and
; TITLE OF INVENTION: Contraception
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1177 West Loop South, 10th Floor
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77027
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,119
; FILING DATE: 09-Oct-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-850-0909
; TELEFAX: 713-850-0165
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..16
; OTHER INFORMATION: /note= "N-Terminal Amino Acid
; Sequence of Pig ZP Glycoprotein II(E)"

```

```
; Patent No. 6264953
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
; US-09-169-119-9

Query Match          35.7%; Score 5; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMGGS 11
Db 12 RMGGS 16

RESULT 5
5177197-32
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 32:
; LENGTH: 41
5177197-32

Query Match          35.7%; Score 5; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 21 GSYRC 25

RESULT 6
US-08-278-089A-25
; Sequence 25, Application US/08278089A
; Patent No. 5681714
; GENERAL INFORMATION:
; APPLICANT: Breitman, Martin L.
; APPLICANT: Rossant, Janet
; APPLICANT: Dumont, Daniel J.
; APPLICANT: Yamaguchi, Terry P.
; TITLE OF INVENTION: No. 5681714el Receptor Tyrosine Kinase
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,089A
; FILING DATE: 20-JUL-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurdydyk, Linda M.
; REGISTRATION NUMBER: 34,971
; REFERENCE/DOCKET NUMBER: 3153-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 25:

; Patent No. 6264953
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
; US-09-169-119-9

Query Match          35.7%; Score 5; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RMGGS 11
Db 12 RMGGS 16

RESULT 5
5177197-32
; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO; OLOFSSON, ANDERS; MOREN, ANITA;
; WERNSTEDT, CHRISTER; HELLMAN, ULF; MIYAZONO, KOHEI; CLAESSON-WELSH,
; LENA; HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/487,343
; FILING DATE: 27-FEB-1990
; SEQ ID NO: 32:
; LENGTH: 41
5177197-32

Query Match          35.7%; Score 5; DB 6; Length 41;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 21 GSYRC 25

RESULT 6
US-08-278-089A-25
; Sequence 25, Application US/08278089A
; Patent No. 5681714
; GENERAL INFORMATION:
; APPLICANT: Breitman, Martin L.
; APPLICANT: Rossant, Janet
; APPLICANT: Dumont, Daniel J.
; APPLICANT: Yamaguchi, Terry P.
; TITLE OF INVENTION: No. 5681714el Receptor Tyrosine Kinase
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,089A
; FILING DATE: 20-JUL-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurdydyk, Linda M.
; REGISTRATION NUMBER: 34,971
; REFERENCE/DOCKET NUMBER: 3153-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 25:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: No. 5681714ch
US-08-278-089A-25

Query Match          35.7%; Score 5; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 28 GSYRC 32

RESULT 7
US-08-838-957A-24
; Sequence 24, Application US/08838957A
; Patent No. 5998187
; GENERAL INFORMATION:
; APPLICANT: Breitman, Martin L.
; APPLICANT: Rossant, Janet
; APPLICANT: Dumont, Daniel J.
; APPLICANT: Yamaguchi, Terry P.
; TITLE OF INVENTION: No. 5998187el Receptor Tyrosine Kinase
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,957A
; FILING DATE: 23-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurdydyk, Linda M.
; REGISTRATION NUMBER: 34,971
; REFERENCE/DOCKET NUMBER: 3153-212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: No. 5998187ch
US-08-838-957A-24

Query Match          35.7%; Score 5; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 28 GSYRC 32
```

RESULT 8
US-08-637-759B-224
; Sequence 224, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 224:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-637-759B-224
Query Match 35.7%; Score 5; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RLPKR 7
Db 13 RLPKR 17
|||||
RESULT 9
US-08-637-759B-224
; Sequence 224, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 224:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-224
Query Match 35.7%; Score 5; DB 3; Length 66;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RLPKR 7
Db 13 RLPKR 17
|||||
RESULT 10
US-09-201-945-224
; Sequence 224, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795

```

; INFORMATION FOR SEQ ID NO: 224:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 66 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
; US-09-201-945-224

Query Match          35.7%; Score 5; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  3 RLPKR 7
DB  13 RLPKR 17

RESULT 11
US-08-264-534-1
; Sequence 1, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,534
; FILING DATE: 03-MAY-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,189
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 7326-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 77 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
; APPLICATION TYPE: peptide
; US-08-264-534-1

Query Match          35.7%; Score 5; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  10 GSYRC 14
DB  23 GSYRC 27

US-08-264-534-1

RESULT 12
US-08-083-590A-14
; Sequence 14, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 77 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: unknown
;   MOLECULE TYPE: peptide
; APPLICATION TYPE: peptide
; US-08-083-590A-14

Query Match          35.7%; Score 5; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  10 GSYRC 14
DB  23 GSYRC 27

US-08-083-590A-14

RESULT 13
US-08-465-500-1
; Sequence 1, Application US/08465500
; Patent No. 5789195
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Muskavitch, Marc A.T.
; APPLICANT: Fehon, Richard G.
; APPLICANT: Rebay, Ilaria
; APPLICANT: Blauumeller, Cristine M.
; APPLICANT: Shepard, Scott B.
; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,500
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-465-500-1

Query Match 35.7%; Score 5; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 23 GSYRC 27

RESULT 14
US-08-346-126-1
; Sequence 1, Application US/08346126
; Patent No. 5849869
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5849869ch And Delta, Binding Domains
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,126
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/791,923
; FILING DATE: 14-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-346-126-1

Query Match 35.7%; Score 5; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 23 GSYRC 27

RESULT 15
US-08-346-128-1
; Sequence 1, Application US/08346128
; Patent No. 5856441
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,128
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/879,038
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-346-128-1

Query Match 35.7%; Score 5; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GSYRC 14
Db 23 GSYRC 27

Search completed: March 4, 2003, 17:02:26
Job time : 7.45763 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 17:00:36 ; Search time 4.50847 Seconds
(without alignments)
130.949 Million cell updates/sec

Title: US-09-745-008-14
Perfect score: 14
Sequence: 1 RORLPKRMGGSYRC 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 188354 seqs, 42170167 residues

Word size : 0
Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	14	10	US-09-745-008-14
2	14	100.0	21	10	US-09-745-008-13
3	14	100.0	1162	10	US-09-745-008-34
4	9	64.3	45	10	US-09-745-008-12
5	6	42.9	332	10	US-09-740-273-2
6	6	42.9	447	9	US-10-086-510-6
7	6	42.9	453	9	US-10-028-072-270
8	6	42.9	453	9	US-10-121-049-270
9	6	42.9	453	9	US-10-123-904-270
10	6	42.9	453	9	US-10-140-470-270
11	6	42.9	453	9	US-10-175-476-270
12	6	42.9	453	9	US-10-176-918-270
13	6	42.9	453	9	US-10-176-921-270
14	6	42.9	453	9	US-10-137-865-270
15	6	42.9	453	9	US-10-140-474-270
16	6	42.9	453	9	US-10-142-431-270
17	6	42.9	453	9	US-10-143-114-270
18	6	42.9	453	9	US-10-140-002-270
19	5	35.7	13	10	US-09-030-619-64

20	5	35.7	15	10	US-09-818-806A-8	Sequence 8, Appli
21	5	35.7	43	10	US-09-864-761-39552	Sequence 39552, A
22	5	35.7	60	10	US-09-864-761-45619	Sequence 45619, A
23	5	35.7	64	10	US-09-754-997A-10	Sequence 10, Appl
24	5	35.7	70	10	US-09-764-877-2048	Sequence 2048, Ap
25	5	35.7	95	10	US-09-764-877-805	Sequence 805, App
26	5	35.7	108	10	US-09-864-761-35069	Sequence 35069, A
27	5	35.7	130	10	US-09-764-853-538	Sequence 538, App
28	5	35.7	143	10	US-09-764-853-730	Sequence 730, App
29	5	35.7	146	10	US-09-815-242-4891	Sequence 4891, Ap
30	5	35.7	146	10	US-09-815-242-5540	Sequence 5540, Ap
31	5	35.7	146	10	US-09-815-242-10534	Sequence 10534, A
32	5	35.7	146	10	US-09-815-242-12225	Sequence 12225, A
33	5	35.7	146	10	US-09-815-242-12826	Sequence 12826, A
34	5	35.7	146	10	US-09-815-242-13049	Sequence 13049, A
35	5	35.7	146	10	US-09-815-242-13217	Sequence 13217, A
36	5	35.7	146	10	US-09-815-242-13575	Sequence 13575, A
37	5	35.7	154	10	US-09-864-761-34493	Sequence 34493, A
38	5	35.7	187	9	US-09-738-626-4821	Sequence 4821, App
39	5	35.7	195	10	US-09-925-301-890	Sequence 890, App
40	5	35.7	203	10	US-09-925-301-992	Sequence 992, App
41	5	35.7	207	9	US-09-895-913A-158	Sequence 158, App
42	5	35.7	211	10	US-09-864-761-36660	Sequence 36660, A
43	5	35.7	230	10	US-09-741-669-372	Sequence 372, App
44	5	35.7	232	9	US-09-738-626-4827	Sequence 4827, Ap
45	5	35.7	242	9	US-09-866-050A-393	Sequence 393, App

ALIGNMENTS

RESULT 1
US-09-745-008-14
; Sequence 14, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-14

Query Match 100.0%; Score 14; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. NO. 2.1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RORLPKRMGGSYRC 14
|||
Db 1 RORLPKRMGGSYRC 14

RESULT 2
US-09-745-008-13
; Sequence 13, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor

; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-13

Query Match 100.0%; Score 14; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14
Db 1 RQRLPKRMGGSYRC 14
|||||

RESULT 3
US-09-745-008-34
; Sequence 34, Application US/09745008
; Patent No. US20020137867A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-745-008-34

Query Match 100.0%; Score 14; DB 10; Length 1162;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYRC 14
Db 379 RQRLPKRMGGSYRC 392
|||||

RESULT 4
US-09-745-008-12
; Sequence 12, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; TITLE OF INVENTION: T. Cruze-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 45

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-09-745-008-12

Query Match 64.3%; Score 9; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KRMGGSYRC 14
Db 30 KRMGGSYRC 38
|||||

RESULT 5
US-09-740-273-2
; Sequence 2, Application US/09740273
; Patent No. US20020042113A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Edmund H. III
; TITLE OF INVENTION: NPRL-Interactors and Methods of Use
; FILE REFERENCE: 1176
; CURRENT APPLICATION NUMBER: US/09/740,273
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/171,691
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Zea mays
US-09-740-273-2

Query Match 42.9%; Score 6; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QRLPKR 7
Db 216 QRLPKR 221
|||||

RESULT 6
US-10-086-510-6
; Sequence 6, Application US/10086510
; Publication No. US20030027258A1
; GENERAL INFORMATION:
; APPLICANT: Fang-Tseh (Frank) CHANG et al.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR PEARL
; TITLE OF INVENTION: OYSTER CULTIVATION
; FILE REFERENCE: 505493000120
; CURRENT APPLICATION NUMBER: US/10/086,510
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 60/310,070
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Pinctada fucata
US-10-086-510-6

Query Match 42.9%; Score 6; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RQRLPK 6
Db 94 RQRLPK 99
|||||

RESULT 7
US-10-028-072-270
; Sequence 270, Application US/10028072
; Publication No. US2003004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063733
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063735
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063738
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064248
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/064809
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065846
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066453
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069212
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069278
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069334
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069694
; PRIOR FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: 60/072320
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 60/073612
; PRIOR FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: 60/074086
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/074092
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081818
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627

; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085149
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 19/98-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 8
US-10-121-049-270
; Sequence 270, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-121-049-270

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 9

US-10-123-904-270
; Sequence 270, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-270

```
Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 10
US-10-140-470-270
; Sequence 270, Application US/10140470
; Publication No. US2003002231A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-270

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 11
US-10-175-746-270
; Sequence 270, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-270

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 12
US-10-176-918-270
; Sequence 270, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-270

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 13
US-10-176-921-270
; Sequence 270, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-270

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RLPKRM 8
Db 43 RLPKRM 48

RESULT 14
US-10-137-865-270
; Sequence 270, Application US/10137865
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-865-270
```

```

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 3 RLPKRM 8
Db 43 RLPKRM 48
```

```

RESULT 15
US-10-140-474-270
; Sequence 270, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 270
; LENGTH: 453
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-474-270
```

```

Query Match 42.9%; Score 6; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 3 RLPKRM 8
Db 43 RLPKRM 48
```

```

Search completed: March 4, 2003, 17:02:58
Job time : 5:50847 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:52:00 ; Search time 6.16949 Seconds
(without alignments)
218.151 Million cell updates/sec

Title: US-09-745-008-14

Perfect score: 78

Sequence: 1 RQLPKRMGGSYRC 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	1162	2 JH0557	exo-alpha-sialidase
2	41	52.6	83	2 AE1851	hypothetical prote
3	41	52.6	584	2 H75403	glycosyl hydrolase
4	40.5	51.9	235	2 J00029	tumor necrosis fac
5	40	51.3	230	2 G65067	hypothetical prote
6	40	51.3	230	2 G91091	hypothetical prote
7	40	51.3	230	2 H85937	hypothetical prote
8	40	51.3	261	2 J00137	hypothetical 30.1K
9	40	51.3	314	2 F75504	probable plin, ty
10	40	51.3	428	2 H86890	hypothetical prote
11	40	51.3	538	2 T51756	NAD+ synthase (glu
12	40	51.3	1559	2 T07757	probable DNA (cyto
13	39	50.0	298	2 B71013	hypothetical prote
14	39	50.0	332	2 G81395	probable ATP /GTP
15	39	50.0	372	2 C90498	carbamate dehydrat
16	39	50.0	395	2 A96215	hypothetical prote
17	39	50.0	395	2 AH3071	conserved hypothet
18	39	50.0	472	2 T50637	probable histidine
19	39	50.0	520	2 S11217	viropilamin - caul
20	39	50.0	1638	2 D87749	protein unc-73b [i
21	39	50.0	1677	2 T14267	kin protein, stage
22	39	50.0	2488	2 T42739	guanine nucleotide
23	38	48.7	210	2 S28673	hypothetical prote
24	38	48.7	356	2 F95954	probable dTPgluco
25	38	48.7	422	2 T21820	hypothetical prote
26	38	48.7	553	2 I37417	hypothetical prote
27	38	48.7	781	2 C69452	glycerol kinase -
28	38	48.7	876	2 T19246	signal-transducing
29	38	48.7	1403	2 S24548	homeotic protein p

30	37	47.4	219	2 T47881	hypothetical prote
31	37	47.4	220	2 E90022	50S ribosomal prot
32	37	47.4	253	2 F64903	transcription regu
33	37	47.4	253	2 H90891	probable ARAC-type
34	37	47.4	253	2 A85726	probable ARAC-type
35	37	47.4	296	2 T47062	hypothetical prote
36	37	47.4	296	2 AC0231	probable binding-p
37	37	47.4	339	2 H86210	hypothetical prote
38	37	47.4	341	2 I61725	natural killer ass
39	37	47.4	347	2 T06671	natural killer cell
40	37	47.4	348	2 A56247	natural killer cell
41	37	47.4	427	2 G02034	killer cell inhibi
42	37	47.4	444	2 G01924	KIR (cl-2) NK rece
43	37	47.4	444	2 G01925	KIR (cl-11) NK rec
44	37	47.4	455	2 G01923	KIR (cl-5) NK rece
45	37	47.4	497	2 T29791	hypothetical prote

ALIGNMENTS

RESULT 1

JH0557
exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi
N;Alternate names: neuraminidase
C;Species: Trypanosoma cruzi
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000
C;Accession: JH0557
R;Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.
J. Exp. Med. 174, 179-191, 1991
A;Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu
A;Reference number: JH0557; MUID:91277609; PMID:1711561
A;Accession: JH0557
A;Molecule type: DNA
A;Residues: 1-1162 <PER>
A;Cross-references: GB:M61732; NID:gl62302; PID:gl62303
A;Note: the authors translated the codon TCR for residue 45 as Cys
C;Comment: This protein plays a role in parasite-host cell interaction.
C;Superfamily: trypanastigote-specific surface antigen
C;Keywords: glycoprotein; glycosidase; hydrolase
F;394/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 2; Length 1162;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RQLPKRMGGSYRC 14
DB 379 RQLPKRMGGSYRC 392

RESULT 2

AE1851
hypothetical protein asr0358 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AE1851
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE1851
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-83 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA072316.1; PID:g17129703; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: asr0358

Query Match 52.6%; Score 41; DB 2; Length 83;

Best Local Similarity 38.5%; Pred. No. 5.4;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Qy 2 QRLPKRMGGSYR 14
: : : : :
Db 46 RNIAKLGGNFR 58

RESULT 3
H75403
glycosyl hydrolase, family 13 - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75403
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75403
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-564 <WHI>
A:Cross-references: GB:AE001983; GB:AE000513; MID:g6459123; PIDN:AAF10944.1; PID:g645912
A:Experimental source: strain.R1
C:Genetics:
A:Gene: DR1375
A:Map position: 1
C:Superfamily: alpha-glucosidase; alpha-amylase core homology

Query Match 52.6%; Score 41; DB 2; Length 564;
Best Local Similarity 61.5%; Pred. No. 33;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RQRLPKRMGGSYR 13
: : : : :
Db 487 RQEQPALVGGSYR 499

RESULT 4
JU0029
tumor necrosis factor alpha precursor - rat
N:Alternate names: cachectin; TNF alpha
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 04-Feb-2000
C:Accession: JU0029; JN0668; S21674
R;Shirai, T.; Shimizu, N.; Horiguchi, S.; Ito, H.
Agric. Biol. Chem. 53, 1733-1736, 1989
A:Title: Cloning and expression in Escherichia coli of the gene for rat tumor necrosis f
A:Reference number: JU0029
A:Accession: JU0029
A:Molecule type: DNA
A:Residues: 1-235 <SHI>
R;Kwon, J.; Chung, I.Y.; Benveniste, E.N.
Gene 132, 227-236, 1993
A:Title: Cloning and sequence analysis of the rat tumor necrosis factor-encoding genes.
A:Reference number: JN0668; MUID:94040766; PMID:8224868
A:Accession: JN0668
A:Molecule type: DNA
A:Residues: 1-235 <KWO>
A:Cross-references: GB:L00981; MID:g205253; PIDN:AAA16275.1; PID:g205254
R;Estler, H.C.; Grewe, M.; Gaussling, R.; Pavlovic, M.; Decker, K.
Biol. Chem. Hoppe-Seyler 373, 271-281, 1992
A:Title: Rat tumor necrosis factor-alpha. Transcription in rat Kupffer cells and in vitr
A:Reference number: S21674; MUID:92329007; PMID:1627266
A:Accession: S21674
A:Molecule type: mRNA
A:Residues: 1-38 'P', 40-162 'T', 164-201 'S', 203-235 <EST>
A:Cross-references: GB:X66539; GB:S40199; MID:g395369; PIDN:CAA47146.1; PID:g395370
C:Comment: Tumor necrosis factor is secreted by macrophages in response to endotoxin and
C:Genetics:
A:Gene: TNF-alpha

A; Introns: 62/3; 81/1; 97/1
C; Superfamily: tumor necrosis factor
C; Keywords: cytokine; cytotoxin; lipoprotein; lymphokine; macrophage; memb
F; 80-235/Product: tumor necrosis factor #status predicted <MAT>
F; 19_20/Binding site: myristate (Lys) (covalent) #status predicted
F; 84/Binding site: carbohydrate (Ser) (covalent) #status predicted
F; 86/Binding site: carbohydrate (Asn) (covalent) #status predicted
F; 148-179/Disulfide bonds: #status predicted

Query Match 51.9%; Score 40.5; DB 2; Length 235;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

Qy 2 QRLPKRMGG---SYRC 14
: |||: ||| | ||
Db 15 EALPKMGGLQNSRRC 30

RESULT 5
G65067
hypothetical protein b2846 - Escherichia coli (strain K-12)
C; Species: Escherichia coli
C; Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C; Accession: G65067
R; Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A; Title: The complete genome sequence of Escherichia coli K-12.
A; Reference number: A64720; MUID: 97426617; PMID: 9278503
A; Accession: G65067
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A; Residues: 1-230 <BLAT>
A; Cross-references: GB:AB000368; GB:U00096; NID:92367165; PIDN:AAC75885.1; PID:gl789212;
A; Experimental source: strain K-12, substrain MG1655

Query Match 51.3%; Score 40; DB 2; Length 230;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRLPKRMGG 10
: |||: ||| ||
Db 207 KQIEKRMGG 216

RESULT 6
G91091
hypothetical protein ECS3703 [imported] - Escherichia coli (strain O157:H7, substrain RIN
C; Species: Escherichia coli
C; Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C; Accession: G91091
R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A; Reference number: A99629; MUID: 21156231; PMID: 11258796
A; Accession: G91091
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-230 <HAY>
A; Cross-references: GB:BA000007; PIDN:BA037126.1; PID:gl3363175; GSPDB:GN00154
A; Experimental source: strain O157:H7, substrain RIMD 0509952
C; Geneticks:
A; Gene: ECS3703

Query Match 51.3%; Score 40; DB 2; Length 230;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRLPKRMGG 10
: |||: ||| ||
Db 207 KQIEKRMGG 216

RESULT 7

B85937
hypothetical protein yqgH [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85937
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85937
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-230 <STO>
A:Cross-references: GB:AE005174; NID:gl2517338; PIDN:AGS7959.1; GSPDB:GN00145; UWGP:Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yqgH

Query Match 51.3%; Score 40; DB 2; Length 230;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RQRLPKRMGG 10

|||:|||||

Db 207 QRLEKRMGG 216

RESULT 8

JQ0137
hypothetical 30.1K protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 07-Jun-1996
C:Accession: JQ0137
R;Kato, J.; Chu, L.; Kitano, K.; DeVault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra, T
Gene 84, 31-38, 1989
A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in P
A:Reference number: JQ0132; MUID:90108714; PMID:2514124
A:Accession: JQ0137
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <KAT>
A:Note: 3-Met could also be the initiator
C:Genetics:
A:Start codon: GTG

Query Match 51.3%; Score 40; DB 2; Length 261;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RQRLPKRMGGSYRC 14

|||||:|||||

Db 31 RQRLRRWGGRRRC 44

RESULT 9

F75504
probable pilin, type IV - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75504
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Shen, M.; Vamathevan, J.O.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-314 <WHI>

Query Match 51.3%; Score 40; DB 2; Length 538;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PKRMGGSYRC 14

|||:|||||

Db 233 PGRIGGSSRC 242

A:Cross-references: GB:AE001913; GB:AE000513; NID:g6458240; PIDN:AAF10127.1; PID:g6458240
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0548
A:Map position: 1

Query Match 51.3%; Score 40; DB 2; Length 314;
Best Local Similarity 72.7%; Pred. No. 28;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RQRLPKRMGG 11

|||||:|||||

Db 26 RQRPQRQGG 36

RESULT 10

H86890
hypothetical protein yvjB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86890
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
A:Reference number: A86625; MUID:21235186; PMID:111337471
A:Accession: H86890
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <STO>
A:Cross-references: GB:AE005176; PID:gl2725185; PIDN:AAK06226.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yvjB

Query Match 51.3%; Score 40; DB 2; Length 428;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PKRMGGSYR 13

|||:|||||

Db 280 PKRMGGSYR 288

RESULT 11

T51756
NAD+ synthase (glutamine-hydrolyzing) (EC 6.3.5.1) [imported] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Nov-2000
C:Accession: T51756
R;Fink, D.; Falke, D.; Wohlleben, W.; Engels, A.
Microbiology 145, 2313-2322, 1999
A:Title: Nitrogen metabolism in Streptomyces coelicolor A3(2): modification of glutamine
A:Reference number: Z25448
A:Accession: T51756
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-538 <FIN>
A:Cross-references: EMBL:Y17736; PIDN:CAB38325.1
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: nade
C:Superfamily: Rhodobacter capsulatus NH3-dependent NAD+ synthase
C:Keywords: ligase

Query Match 51.3%; Score 40; DB 2; Length 538;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 PKRMGGSYRC 14

|||:|||||

Db 233 PGRIGGSSRC 242

RESULT 12

T07757
probable DNA (cytosine-5-)-methyltransferase (EC 2.1.1.37) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 05-May-2000
C:Accession: T07757
R:Cella, R.

submitted to the EMBL Data Library, October 1997
A:Reference number: Z16119

A:Accession: T07757
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1559 <CEL>

A:Cross-references: EMBL:AJ002140; NID:e1250983; PIDN:CAA05207.1; PID:e1250984

A:Experimental source: cultivar Meraviglia; apical meristems

C:Genetics:

A:Gene: SMT

C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 51.3%; Score 40; DB 2; Length 1559;

Best Local Similarity 59.3%; Pred. No. 1.3e+02;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 13

||| ||| :|

Db 1423 QRIPKPGADWR 1434

RESULT 13

B71013

hypothetical protein PH1400 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999

C:Accession: B71013

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamanoto, S.; Sekin
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: B71013

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-298 <KAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30506.1; PID:d1031449; PID:g32578

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1400

Query Match 50.0%; Score 39; DB 2; Length 298;

Best Local Similarity 63.6%; Pred. No. 40;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 RLKPKRMGGSYR 13

||| ||| |

Db 51 RLKRLGGIYK 61

RESULT 14

G81395

probable ATP /GTP binding protein Cj0500 [imported] - Campylobacter jejuni (strain NCTC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: G81395

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: G81395

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <PAR>

A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75138.1; PID:g6967971

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0500

Query Match 50.0%; Score 39; DB 2; Length 332;

Best Local Similarity 61.5%; Pred. No. 45;

Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRLPKRMGGSYR 14

||| ||| |||

Db 57 QALAKAMGASTYC 69

RESULT 15

C90498

carnitine dehydratase, probable [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 01-Mar-2002

C:Accession: C90498

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-v
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: C90498

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE006641; NID:gl3816568; PIDN:AAK43242.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO3144

C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv3272

Query Match 50.0%; Score 39; DB 2; Length 372;

Best Local Similarity 54.5%; Pred. No. 50;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 RLKPKRMGGSYR 13

||| ||| |||

Db 209 KVPKRMGSAHR 219

Search completed: March 4, 2003, 16:56:56

Job time : 8.16949 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2003, 16:52:00 ; Search time 19.8305 Seconds
(without alignments)
218.151 Million cell updates/sec

Title: US-09-745-008-12

Perfect score: 229

Sequence: 1 QPLRRQVVVVPLSPRLVLL.....RLPLKRMGGSYRCVNASTAN 45

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73:*

1: piri:*

2: piri:*

3: piri:*

4: piri:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	214.5	93.7	1162	2 JH0557	exo-alpha-sialidas
2	59.5	26.0	716	2 AG2493	hypothetical prote
3	59.5	26.0	1148	2 D83270	transcription-repa
4	59	25.8	94	2 T03751	hypothetical prote
5	58.5	25.5	216	2 S40720	hypothetical prote
6	58	25.3	608	2 AG0846	formate hydrogenly
7	56.5	24.7	347	2 T51478	root cap 1 (RCp1)
8	56.5	24.7	677	2 AH2573	hypothetical prote
9	55	24.0	608	2 G65052	hydrogenase (EC 1.
10	55	24.0	608	2 C91076	membrane-spanning
11	55	24.0	608	2 B85921	hypothetical prote
12	55	24.0	651	2 AF2493	hypothetical prote
13	53.5	23.4	304	2 F83229	probable short cha
14	52.5	22.9	247	2 T16770	hypothetical prote
15	51	22.3	500	1 E69291	4-hydroxyphenylace
16	51	22.3	629	2 F84428	probable myosin he
17	51	22.3	686	2 S61621	probable membrane
18	50.5	22.1	82	2 H87312	hypothetical prote
19	50.5	22.1	177	2 A54377	interleukin-1 rece
20	50.5	22.1	234	2 A70736	hypothetical prote
21	50.5	22.1	302	2 S21357	beta-casein - tam
22	50	21.8	407	2 C82327	MSHA biogenesis pr
23	50	21.8	835	2 T42970	primase - ateline
24	50	21.8	879	2 S23006	shed acute-phase a
25	50	21.8	1819	2 E86410	protein F3M18.14
26	49.5	21.6	268	2 H91087	probable enzyme
27	49.5	21.6	268	2 B85933	probable enzyme yg
28	49.5	21.6	268	2 H65063	hypothetical prote
29	49.5	21.6	305	2 B64960	membrane protein y

hypothetical prote
hypothetical prote
conserved hypothet
conserved hypothet
groovin gene prote
5-formyltetrahydro
Mg protoporphyrin
hypothetical prote
conserved hypothet
conserved hypothet
K0885.3 protein -
conserved hypothet
oxidoreductase, pr
hypothetical prote
conserved hypothet
probable OmpA fami

ALIGNMENTS

RESULT 1

JH0557

N;Alternate names: neuraminidase

C;Species: Trypanosoma cruzi

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000

C;Accession: JH0557

R;Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.

J. Exp. Med. 174, 179-191, 1991

A;Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu

A;Reference number: JH0557; MUID:91277609; PMID:1711561

A;Accession: JH0557

A;Molecule type: DNA

A;Residues: 1-1162 <PER>

A;Cross-references: GB:M61732; NID:G162302; PID:G162303

A;Note: the authors translated the codon TCT for residue 45 as Cys

C;Comment: This protein plays a role in parasite-host cell interaction.

C;Superfamily: trypanastigote-specific surface antigen

C;Keywords: glycoprotein; glycosidase; hydrolase

F;394/Binding site: carbohydrate (Asn) #status predicted

Query Match 93.7%; Score 214.5; DB 2; Length 1162;

Best Local Similarity 97.8%; Pred. NO. 1.1e-21;

Matches 44; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QPLRRQVVVVPLSPRLVLLAFRCRLPLKRMGGSYRCVNASTAN 45

DB 356 QPLRRQVVVVPLSPRLVLLAFRCRLPLKRMGGSYRCVNASTAN 399

RESULT 2

AG2493

hypothetical protein alr7127 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alr

C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C;Accession: AG2493

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG2493

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-716 <KUR>

A;Cross-references: GB:BA000020; PIDN:BA078211.1; PID:G1713565; GSPDB:GN00180

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr7127

A;Genome: plasmid

```

Query Match          26.0%; Score 59.5; DB 2; Length 716;
Best Local Similarity 29.8%; Pred. No. 3.6;
Matches 14; Conservative 10; Mismatches 10; Indels 13; Gaps 1;

QY      7 RVVVPLSPRLVL-----LAFCRQLPLKRMGGSYRCVN 40
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      655 RVCIEHFTLRLLLYCNGYIPWYARFLDYCTERMLLQRVGGYRFFIH 701
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 3
D83270
transcription-repair coupling protein Mfd PA3002 [imported] - Pseudomonas aeruginosa (st
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83270
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83270
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1148 <STO>
A:Cross-references: GB:AE004725; GB:AE004091; NID:g9949100; PIDN:AAG06390.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mfd; PA3002
C:Superfamily: transcription-repair coupling protein

Query Match          26.0%; Score 59.5; DB 2; Length 1148;
Best Local Similarity 33.9%; Pred. No. 5.6;
Matches 19; Conservative 7; Mismatches 13; Indels 17; Gaps 2;

QY      2 PLRQRVVVPLSPRLVLLAFCR-----QRLPLKRM-----GGSYRCVN 40
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      99 POLKHGVLVPISTALHRLAPTRFLGSSVLVDVGQKLDVEMRLLEGAGYRCVD 154
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 4
T03751
hypothetical protein A - slime mold (Physarum polycephalum) mitochondrion integrated pla
C:Species: mitochondrion Physarum polycephalum
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: T03751
R:Nakagawa, C.C.; Jones, E.P.; Miller, D.L.
Curr. Genet. 33, 178-187, 1998
A:Title: Mitochondrial DNA rearrangements associated with mF plasmid integration and pla
A:Reference number: Z15055; MUID:98177147; PMID:9508792
A:Accession: T03751
A>Status: preliminary; translated from GB/EMBL/DBDB
A:Molecule type: DNA
A:Residues: 1-94 <NAK>
A:Cross-references: EMBL:AF012249; NID:g2435419; PIDN:AAC15931.1; PID:g2435420
A:Experimental source: strain aux2-S
C:Genetics:
A:Genome: mitochondrion integrated plasmid mF
C:Keywords: mitochondrion

Query Match          25.8%; Score 59; DB 2; Length 94;
Best Local Similarity 30.9%; Pred. No. 0.59;
Matches 17; Conservative 11; Mismatches 15; Indels 12; Gaps 2;

QY      3 LRRQRVVVPLSPRLVLLAF-CRORLPLKRM-----GGSYRCVNASTAN 45
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      15 LKRNFLVFPPTISLLISWCKRALLIKLIYSIHVILFIKNTSIRCYSTNSN 69
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 5
S40720
hypothetical protein C48B4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

```

Query Match	24.0%;	Score 55;	DB 2;	Length 608;
Best Local Similarity	30.8%;	Pred. No. 13;		
Matches	12;	Conservative	10;	Mismatches 15; Indels 2; Gaps 2;
QY	1	QPLRRQRRVVVPLSPRLVLLAFCR-QRLPLKRMGGSYRC	38	
DB	500	QPMITLLLIACPLLP-FIIMAICKGDRLPSSRGAAWVC	537	
<p>RESULT 10</p> <p>C91076</p> <p>membrane-spanning protein of hydrogenase 3 [imported] - Escherichia coli (strain O157:H7, C:Species: Escherichia coli</p> <p>C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001</p> <p>C/Accession: C91076</p> <p>R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; gasawaka, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.</p> <p>DNA Res. 8, 11-22, 2001</p> <p>A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome A:Reference number: A99629; MUID:21156231; PMID:11258796</p> <p>A/Accession: C91076</p> <p>A>Status: preliminary</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-608 <HAY></p> <p>A/Cross-references: GB:BA000007; PIDN:BAB37002.1; PID:gl3363050; GSPDB:GN00154</p> <p>A/Experimental source: strain O157:H7, substrain RIMD 0509952</p> <p>C/Genetics:</p> <p>A:Gene: ECs3579</p> <p>C:Superfamily: formate hydrogenlyase chain 3</p>				
Query Match	24.0%;	Score 55;	DB 2;	Length 608;
Best Local Similarity	30.8%;	Pred. No. 13;		
Matches	12;	Conservative	10;	Mismatches 15; Indels 2; Gaps 2;
QY	1	QPLRRQRRVVVPLSPRLVLLAFCR-QRLPLKRMGGSYRC	38	
DB	500	QPMITLLLIACPLLP-FIIMAICKGDRLPSSRGAAWVC	537	
<p>RESULT 11</p> <p>B85921</p> <p>hypothetical protein hycC [imported] - Escherichia coli (strain O157:H7, substrain EDL93: C:Species: Escherichia coli</p> <p>C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001</p> <p>C/Accession: B85921</p> <p>R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dialanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001</p> <p>A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.</p> <p>A/Reference number: A85480; MUID:21074935; PMID:11206551</p> <p>A/Accession: B85921</p> <p>A>Status: preliminary</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-608 <STO></p> <p>A/Cross-references: GB:AE005174; NID:gl215171; PIDN:AAG57830.1; GSPDB:GN00145; UWGP:Z40: A/Experimental source: strain O157:H7, substrain EDL933</p> <p>C/Genetics:</p> <p>A:Gene: hycC</p> <p>C:Superfamily: formate hydrogenlyase chain 3</p>				
Query Match	24.0%;	Score 55;	DB 2;	Length 608;
Best Local Similarity	30.8%;	Pred. No. 13;		
Matches	12;	Conservative	10;	Mismatches 15; Indels 2; Gaps 2;
QY	1	QPLRRQRRVVVPLSPRLVLLAFCR-QRLPLKRMGGSYRC	38	
DB	500	QPMITLLLIACPLLP-FIIMAICKGDRLPSSRGAAWVC	537	
<p>RESULT 12</p> <p>AF2493</p> <p>hypothetical protein alr7126 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1; C:Species: Nostoc sp.</p>				

